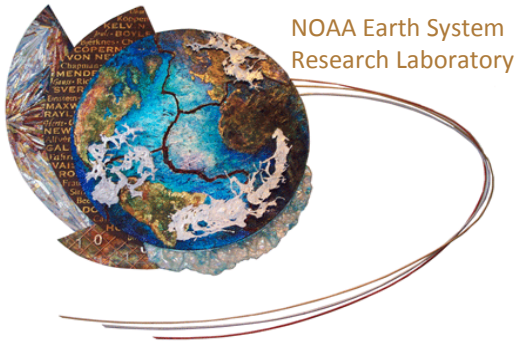


TIGGE multi-model and ECMWF reforecast-calibrated probabilistic precipitation forecasts over the US

Tom Hamill
NOAA ESRL, Physical Sciences Division
tom.hamill@noaa.gov
(303) 497-3060

For more information

- MWR article, just submitted:
  - **http://tinyurl.com/TIGGE-ref-pdf**
- Online appendix, with more complete set of figures:
  - **http://tinyurl.com/TIGGE-ref-app**

Treating "system error" in ensembles

- System error includes errors due to *model imperfections* as well as *sampling error* from finite ensemble.
- May manifest itself as biased mean, under-spread, poorly forecast higher moments.
- Treat through:
  - higher resolution & other model improvements
  - (physically based) stochastic parameterizations
  - multi-model
  - statistical post-processing

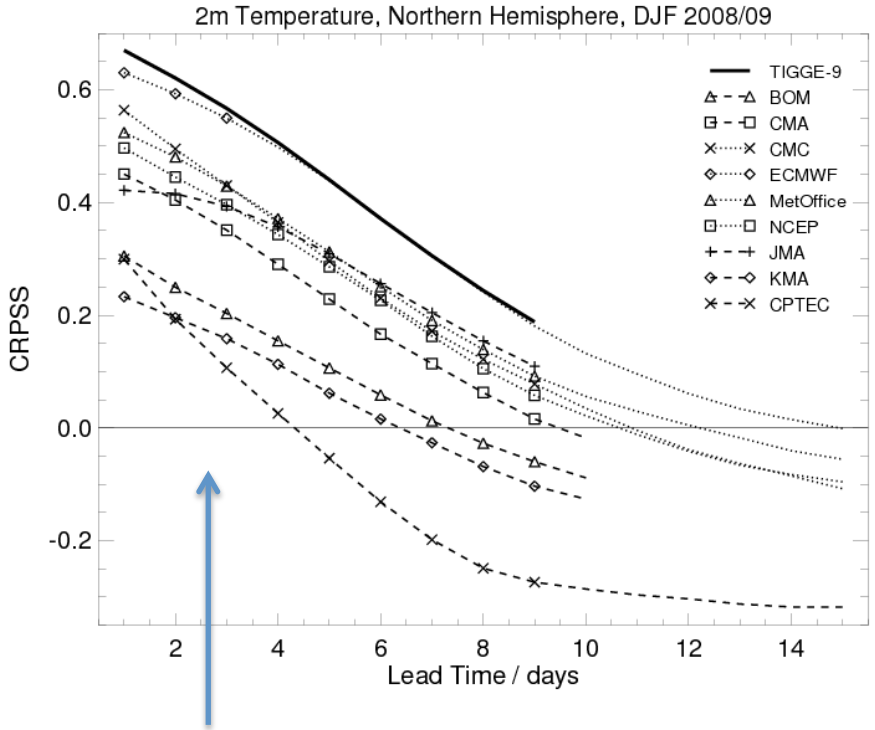
Reforecast-based statistical post-processing

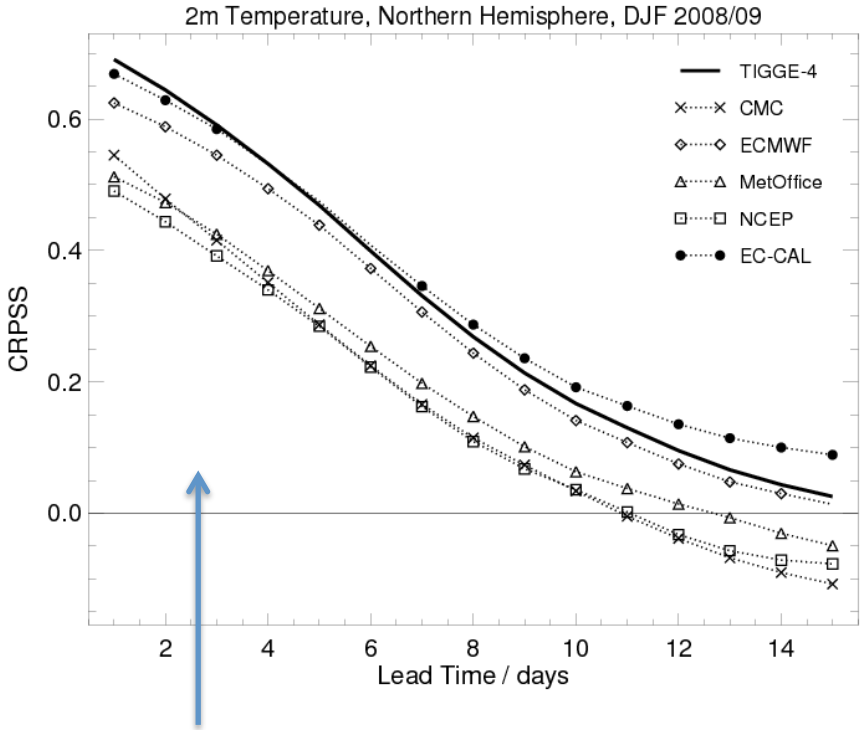
- Advantages:
  - Can ameliorate systematic errors for tough forecast problems, such as long-lead forecasts, rare events.
  - Can provide highly reliable ensemble guidance, improving user confidence.
- Disadvantages:
  - Best results with long training data set (i.e., reforecasts), which are computationally expensive to compute, and
  - This makes NWP centers less willing to rapidly change the model (else reforecasts differ from real-time forecast).

Perceived multi-model advantages and disadvantages

- Advantages:
  - Basically free information; only need to have the extra bandwidth, storage ready.
  - If modeling systems relatively independent, then some cancellation of errors, improved diversity of forecasts.
- Disadvantages:
  - Not all centers yet willing to share their data in real time.
  - Creates dependencies on other centers. Can they provide their data in accordance with your timelines?
  - Multiple systems that can change, not just your own. Rarely stable for long.
  - MM concept and results may not be generalizable; it may matter specifically which models are used. Combinations of immature models may not provide much improvement.

Multi-model combination: better than the best model?



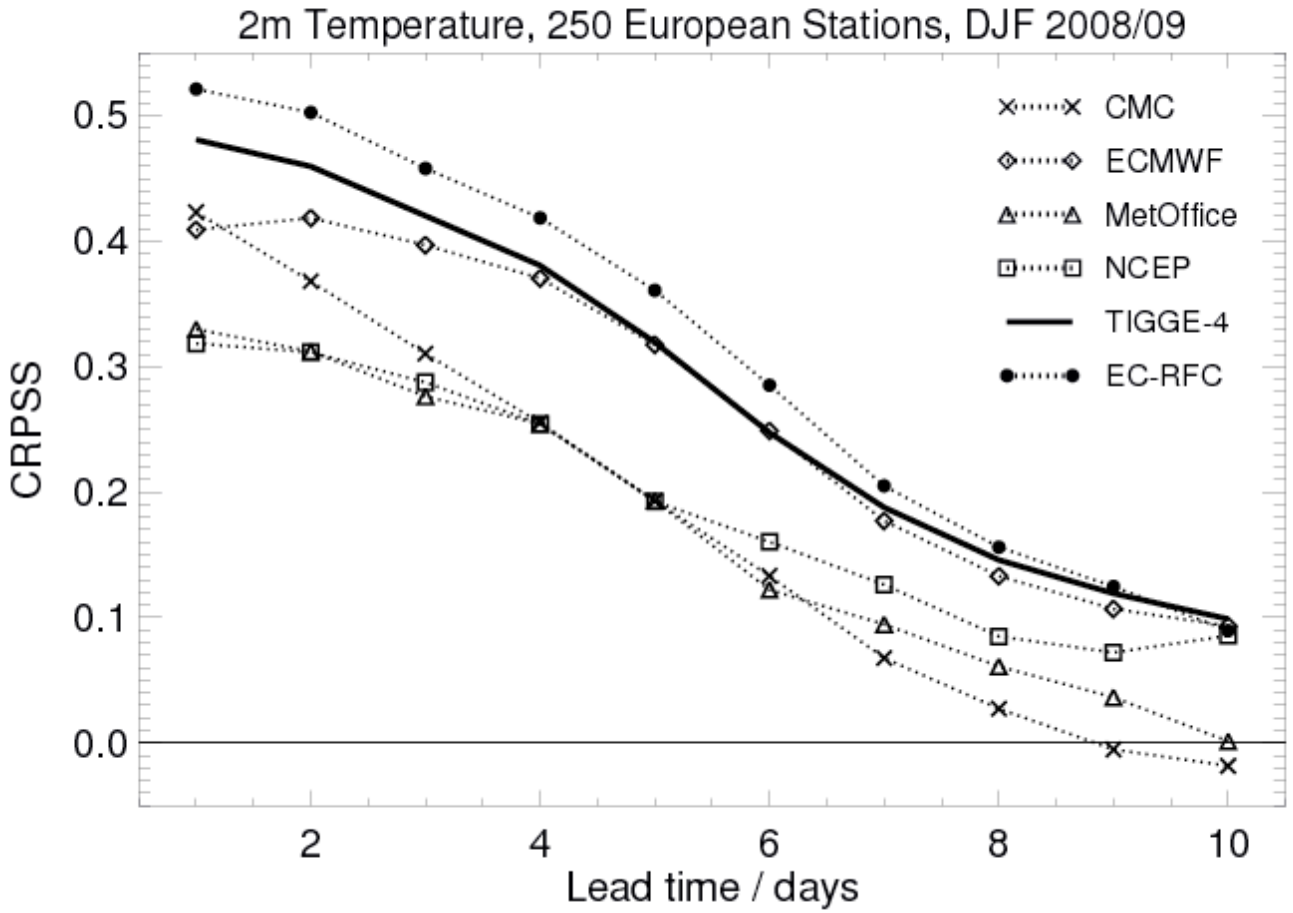


9-model MM ensemble little better than the best system, ECMWF. Forecasts are bias-corrected using last 30 days of F-A differences. ECMWF analysis used as reference (somewhat problematic).

4-model MM ensemble better than the best system, ECMWF. Poorer performing ensemble systems drag down the MM performance. Also: reforecast-calibrated ECMWF competitive

courtesy of Renate Hagedorn, ECMWF & DWD. Hagedorn et al., QJRMS, submitted.

Previously, reforecast vs. multi-model, T_{sfc}



ECMWF's forecasts were corrected here using a blend of bias correction from the past 30 days of forecasts and a more sophisticated regression approach using reforecasts.

courtesy of Renate Hagedorn, ECMWF & DWD. Hagedorn et al., *QJRMS*, submitted.

Hypothesis

- Reforecast-based calibration of daily precipitation, like T_{sfc}, will be more skillful than multi-model forecasts of precipitation, with or without multi-model calibration using a short training data set.

Probabilistic forecasts to be compared (perturbed members, no control) Jul – Oct 2010 over CONUS, 00Z only

- NCEP operational, 20 members, T190L28.
- ECMWF operational, first 20 members, T639L62.
- UK Met Office, 20 members.
- CMC, 20 members.
- Multi-model (80 members).
- Multi-model calibrated using prior 30 days of forecasts/ analyses (more detail later).
- ECMWF with reforecast calibration (more detail later).

Precipitation verification data set

- Use NCEP/EMC "CCPA" dataset of Stage-IV precipitation, regression-corrected to CPC analysis over CONUS, and upscaled to 1-degree. Described in Luo et al. (2010).
  - some points in western US where regression correction fails due to lack of data. Substitute upscaled Stage-IV for those points
- Verify only where 1-degree box is within CONUS (conterminous US).
- Verify 24-h accumulations.
- Verify dates from 00Z 1 July 2010 to 00Z 31 Oct 2010.

ECMWF's reforecast data set

- Once weekly (1 Jan, 8 Jan, 15 Jan, etc.) 5-member ensemble, control + 4 perturbed.
- Past 18 years; we use only 2002-2009
- Control = ERA-Interim reanalysis (using slightly out-of-date 4D-Var and older forecast model version).
- Perturbed from combination of 2010's perturbed-obs 4D-Var perturbations (not flow dependent) + singular vectors appropriate to the past date.
- Uses same forecast model as is used for operational EPS system.

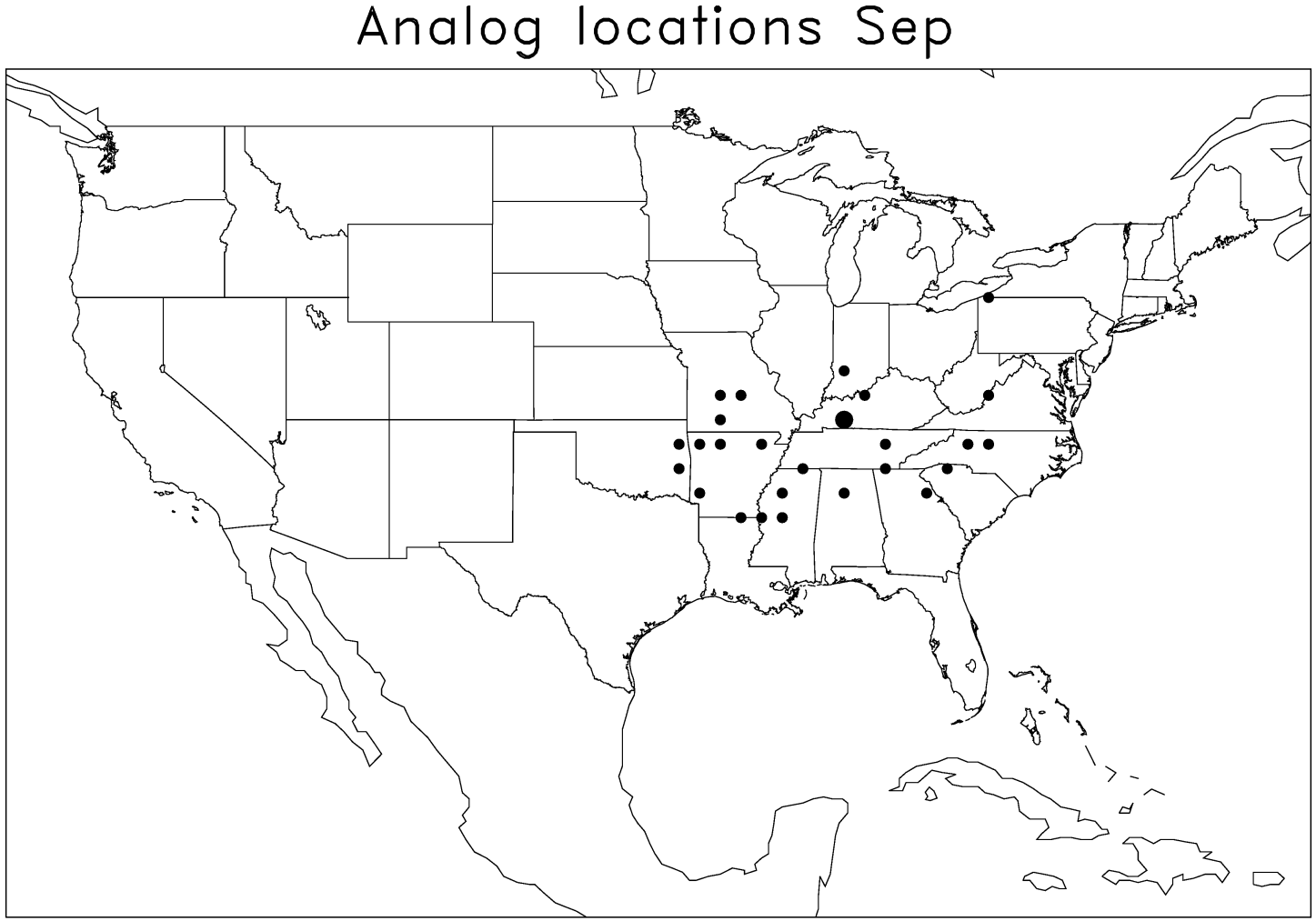
Post-processing method: "extended" logistic regression

- Follows Wilks' (2009, *Met Apps*) approach to provide full probability distribution.

$$P(obs > T) = \frac{\exp\left[b_0 + b_1 \overline{x}^{0.4} + b_1 \overline{x}^{0.4} \sigma^{0.8} + b_2 T^{0.4} \right]}{1 + \exp\left[b_0 + b_1 \overline{x}^{0.4} + b_1 \overline{x}^{0.4} \sigma^{0.8} + b_2 T^{0.4} \right]}$$

- Train on 2002-2009 data only, since that's the period when precipitation analyses available
  - Thus, can't fully use 18-year ECMWF weekly 5-member reforecast.
  - Augment samples using 25 nearby grid points with similar analysis CDFs, especially at the high amounts.

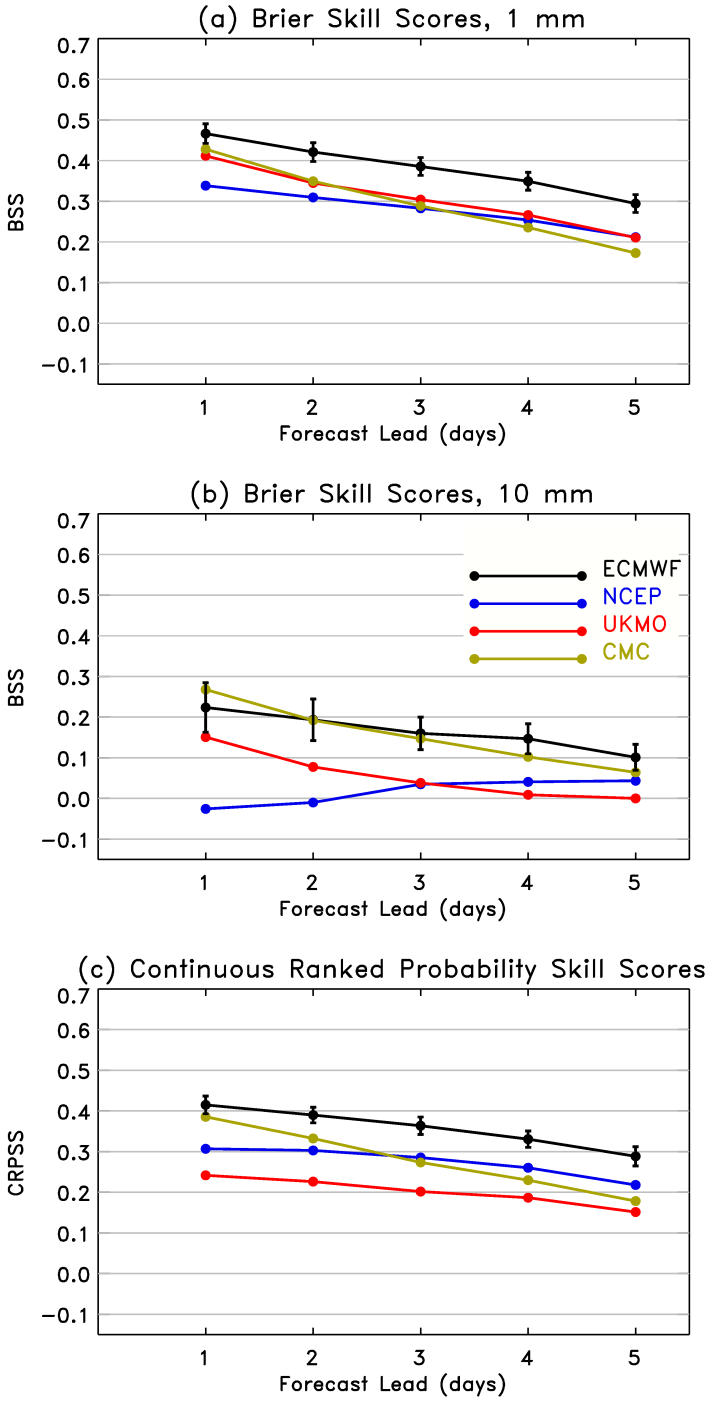
Example of supplementary training data locations



Primary verification techniques

- Brier skill scores and CRPSS
  - Calculated in manner to avoid skill overestimate following Hamill and Juras, *QJRMS*, Oct 2006.
  - Details in supplementary slides and in article, http://tinyurl.com/TIGGE-ref-pdf
- Reliability diagrams
- Confidence intervals via paired block bootstrap following Hamill, *WAF*, 1999.

Results directly from ensemble systems, no post-processing, no multi-model



Skill scores of various 20-member ensembles

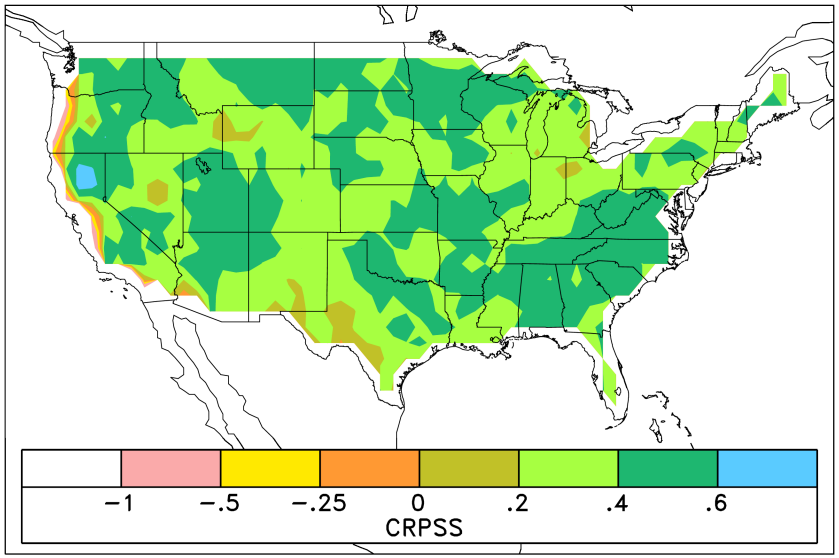
5th and 95th percentiles using block bootstrap algorithm following Hamill, WAF, 1999.

ECMWF generally the most skillful, though CMC makes similarly skillful 10-mm forecasts.

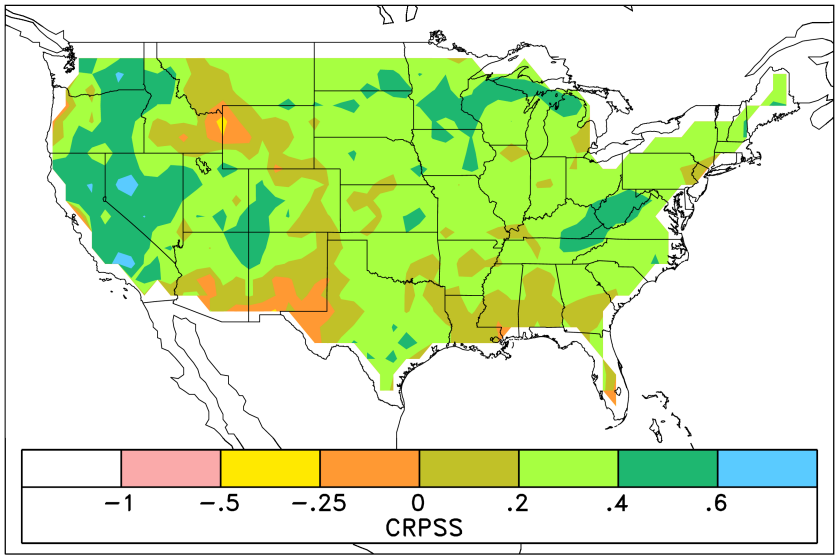
NCEP and UKMO trail.

CRPSS geographical distributions

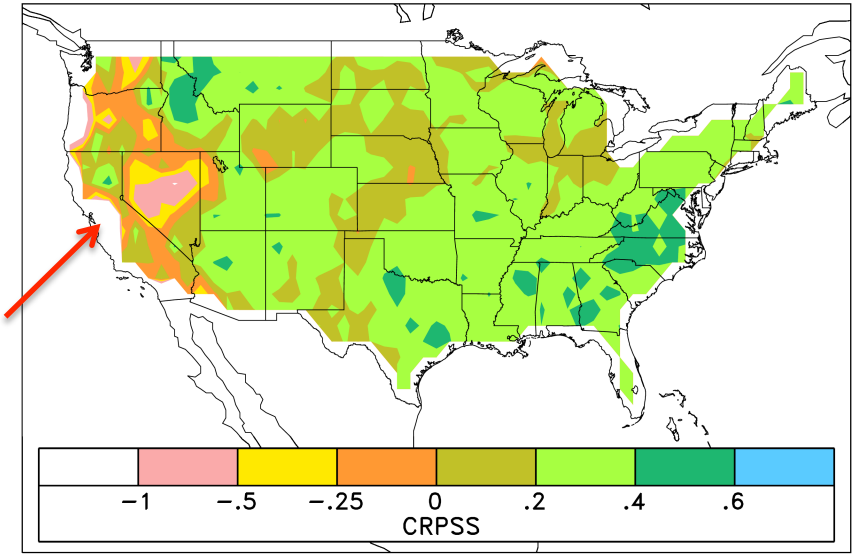
(a) ECMWF CRPSS Day +3



(b) NCEP CRPSS Day +3

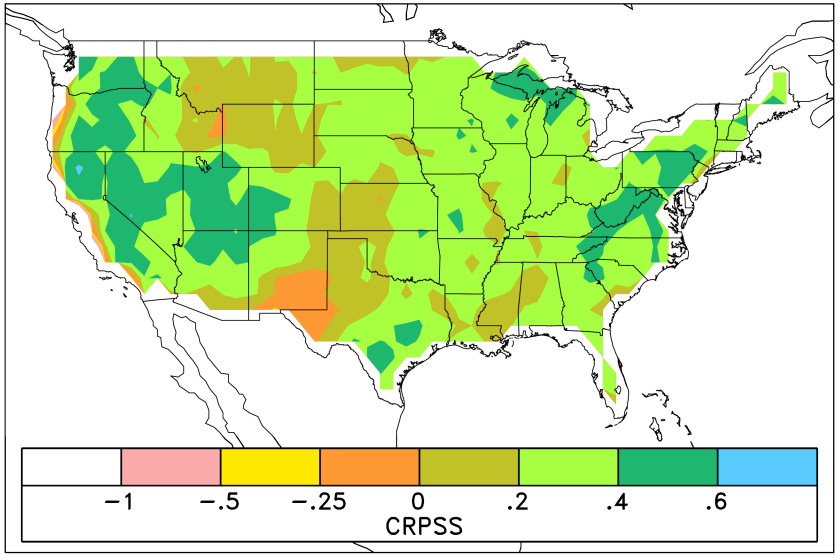


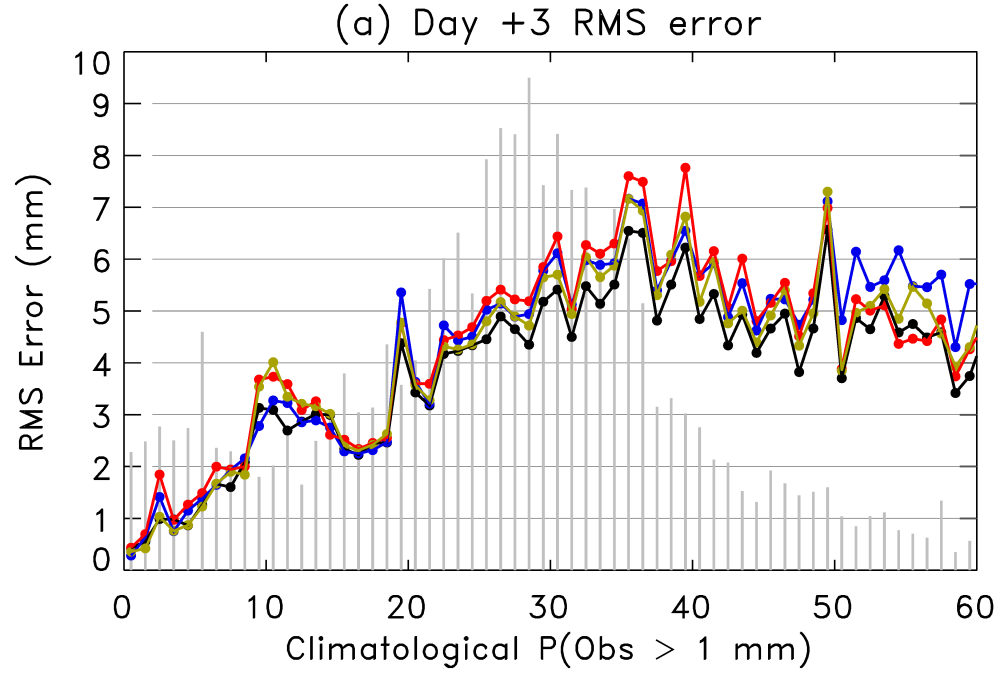
(c) UKMO CRPSS Day +3

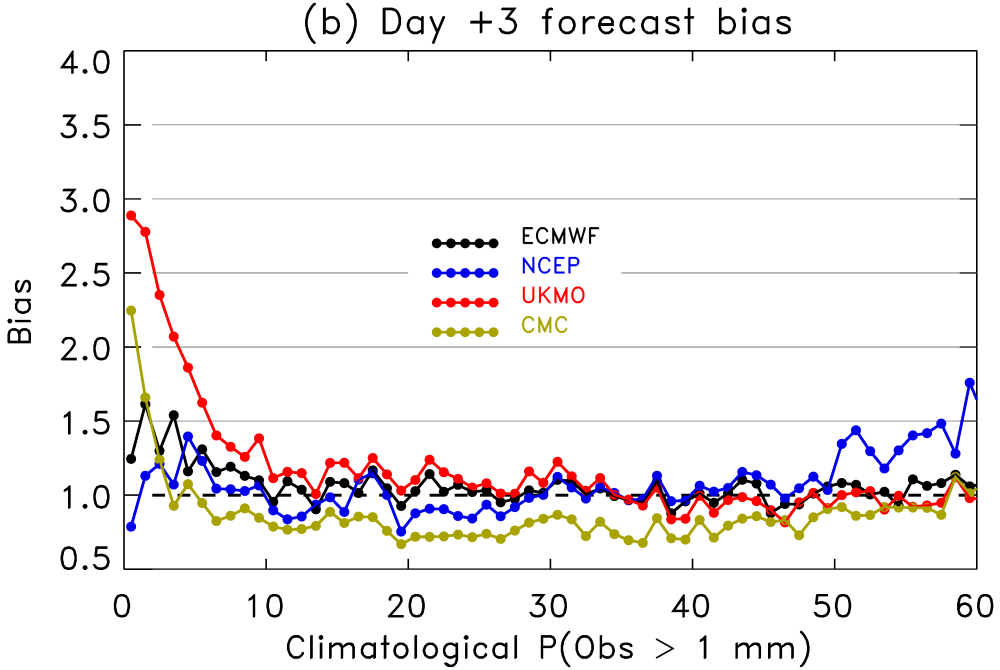


?

(d) CMC CRPSS Day +3



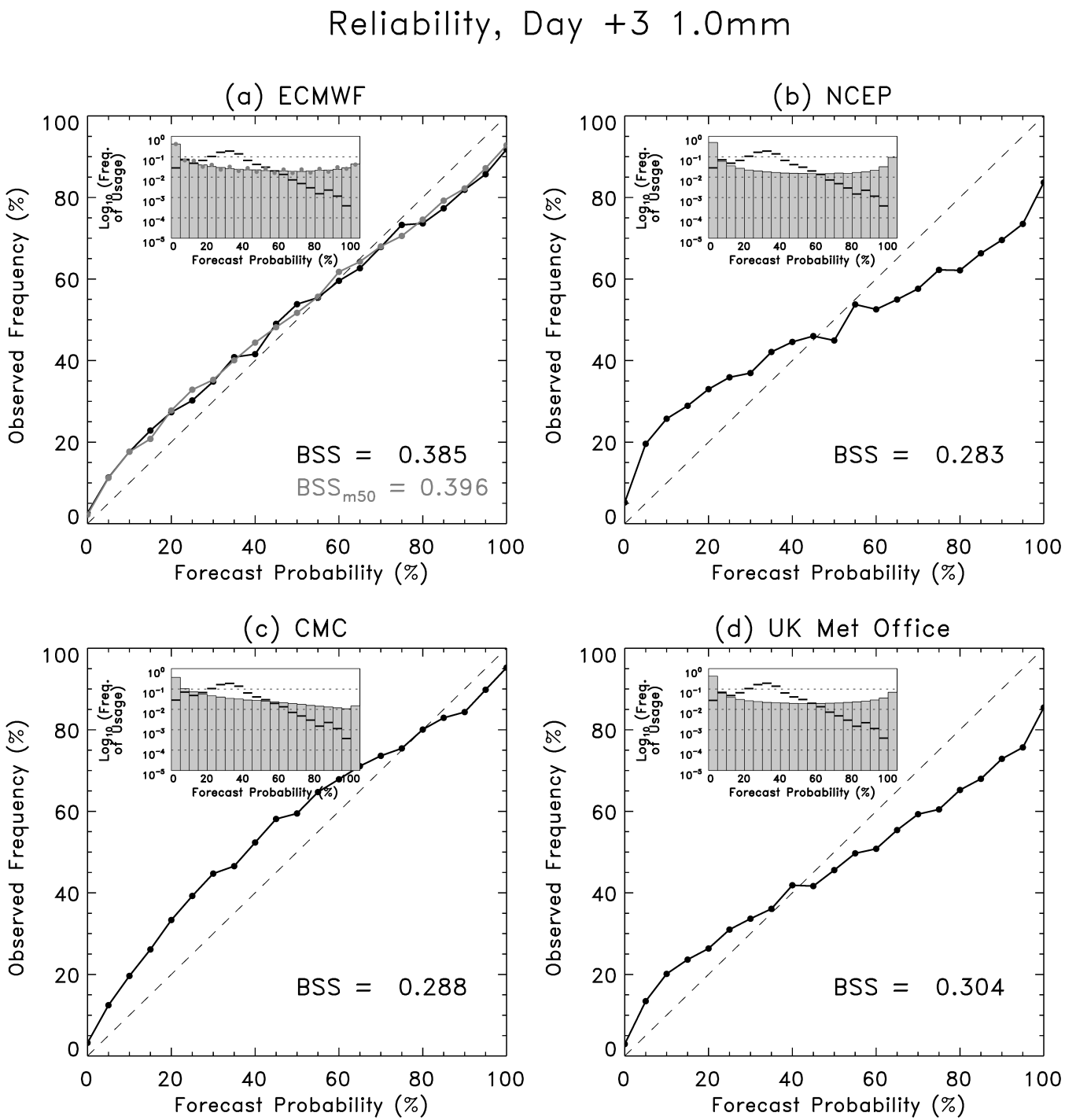




RMS error and bias as f(climatological probability)

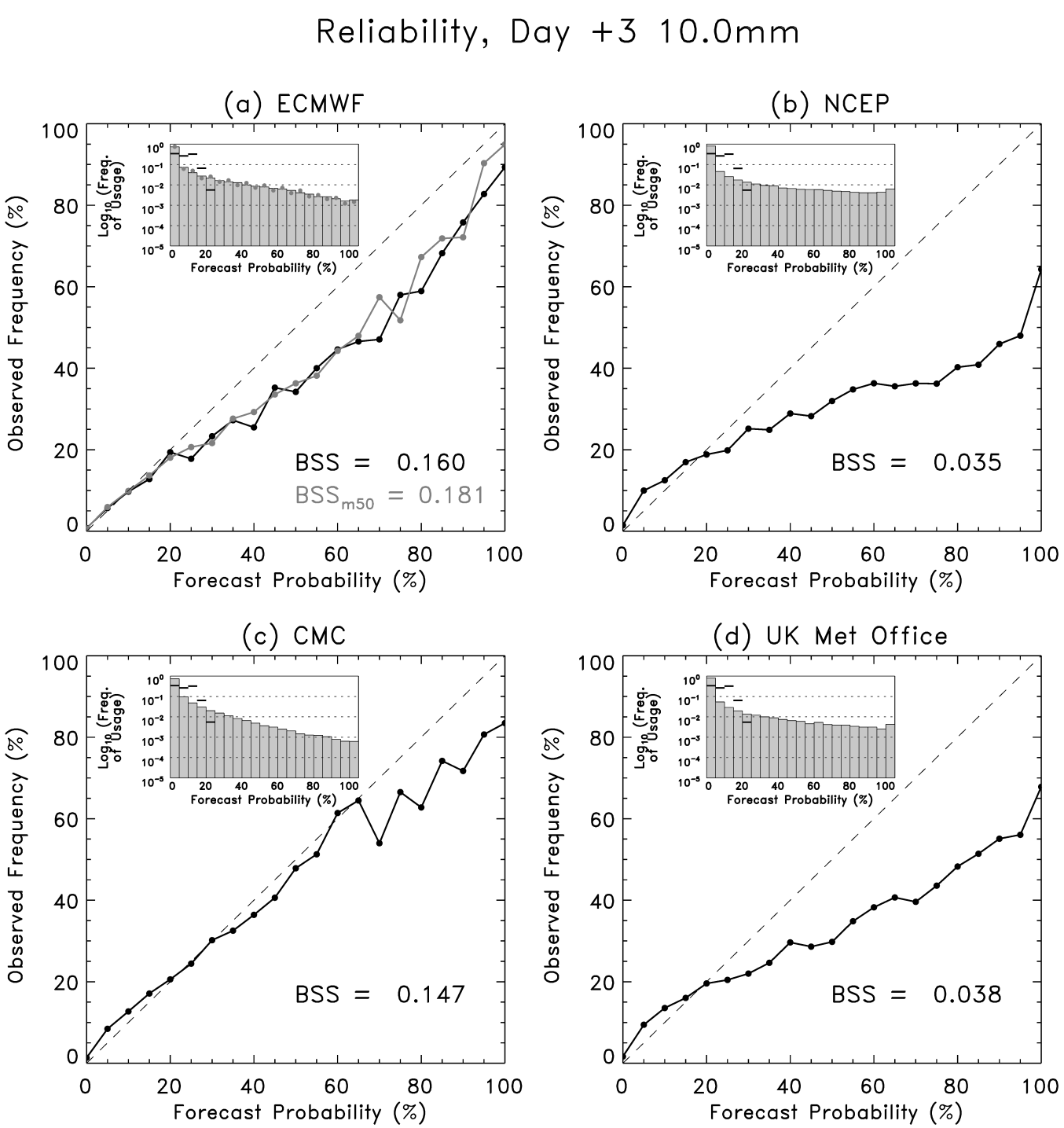
Errors tend to be small when probabilities are small; likely most observed events are light precipitation.

Note large over-forecast bias of UKMO for the climatologically dry areas. This is responsible for UKMO's negative CRPSS in dry regions.

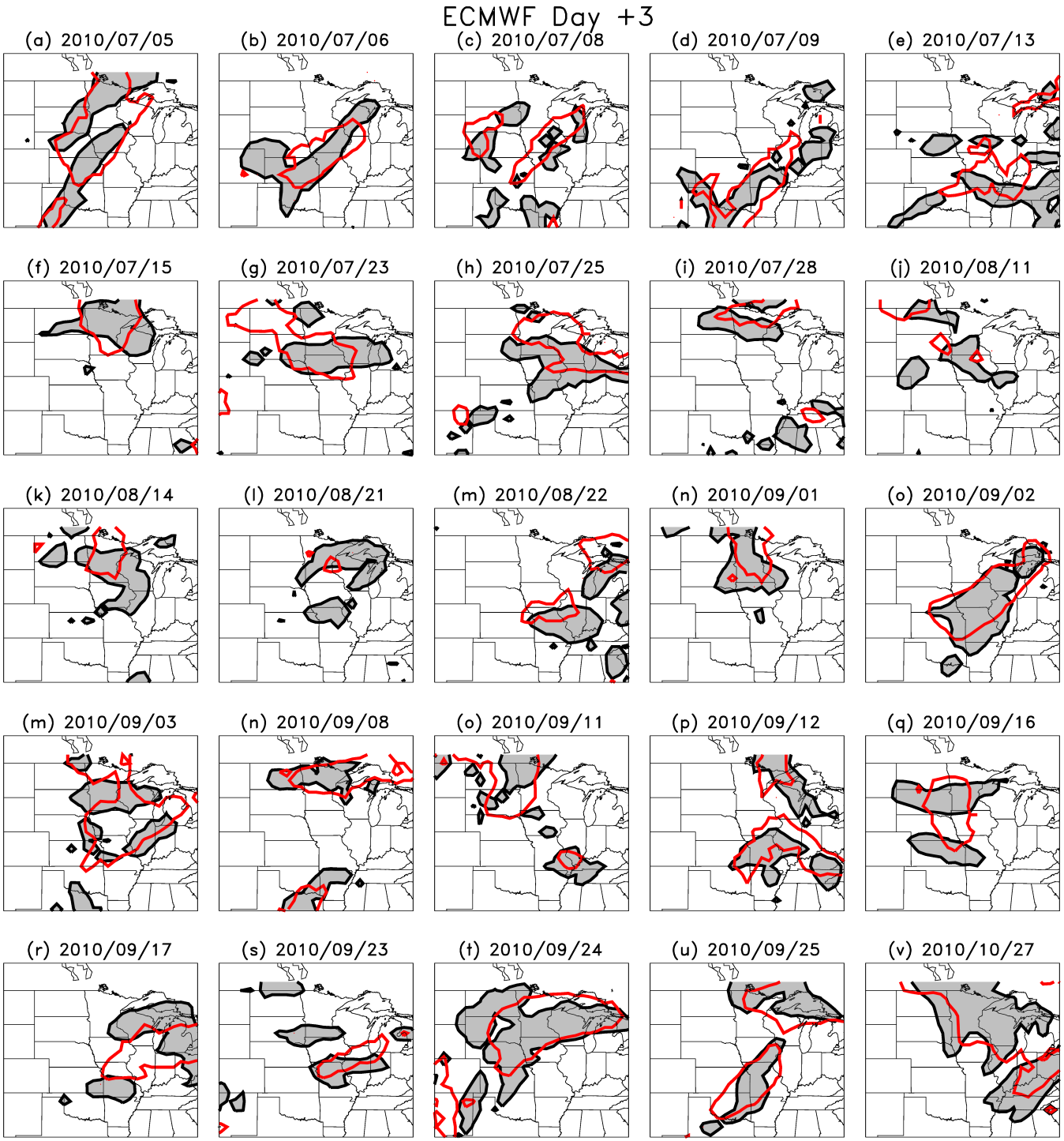


Reliability diagrams, day +3

> 1.0 mm



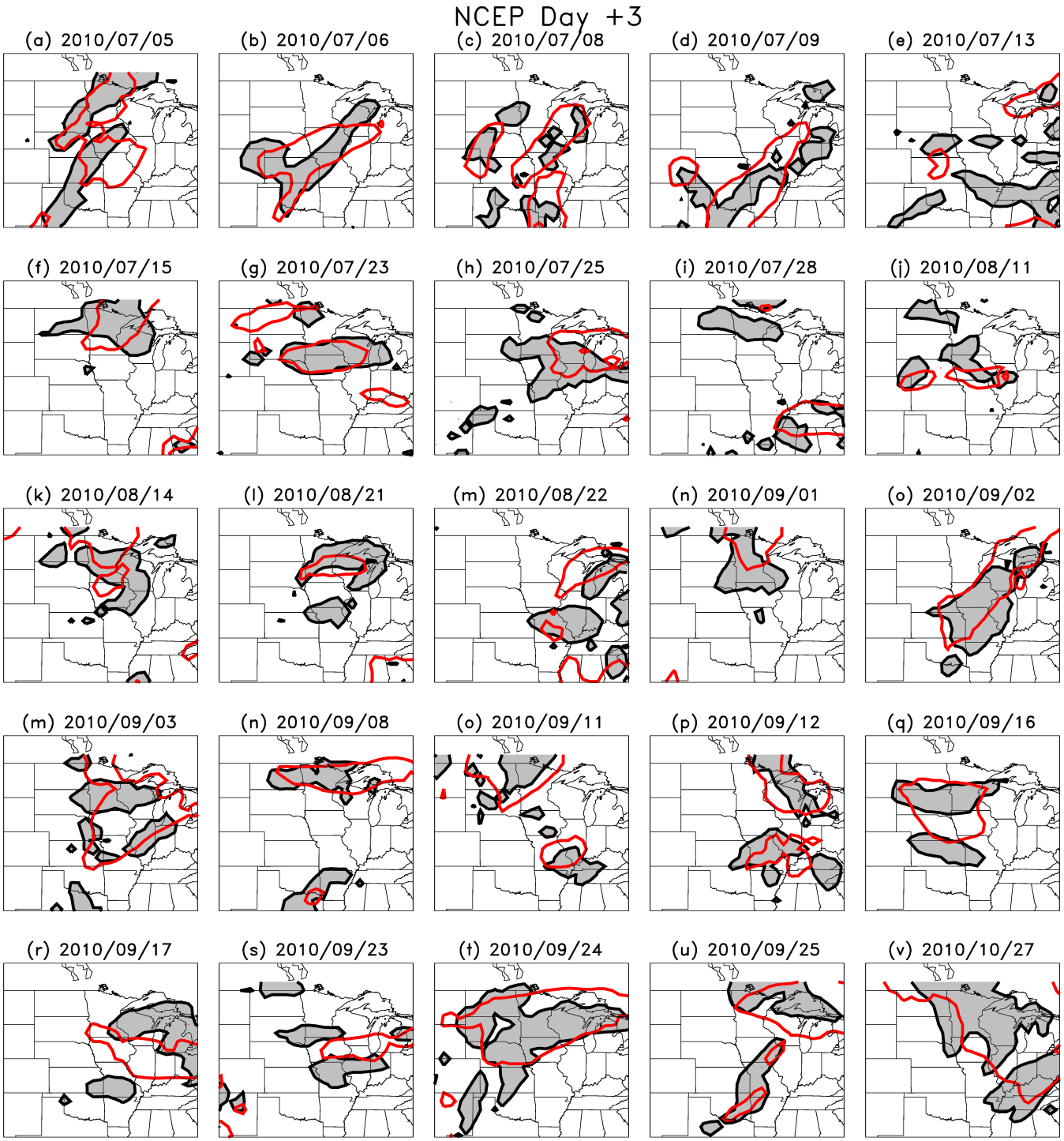
Reliability diagrams, day +3, > 10 mm



Positional biases?

Black: analyzed > 10-mm 24 h^{-1} area.

Red: > 50% forecast contour for 10-mm 24 h^{-1} area.

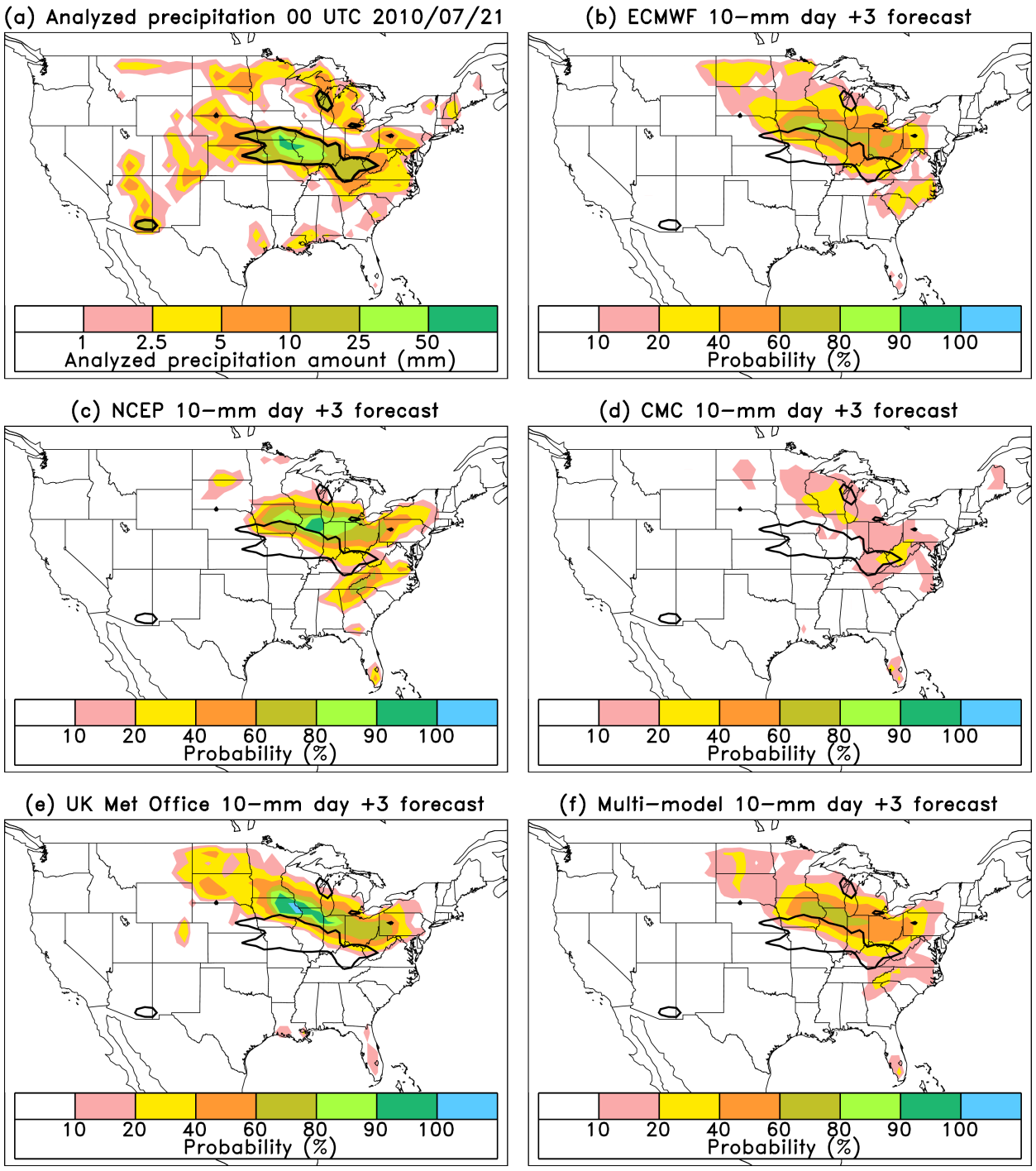


Positional biases?

Black: analyzed > 10-mm 24 h^{-1} area.

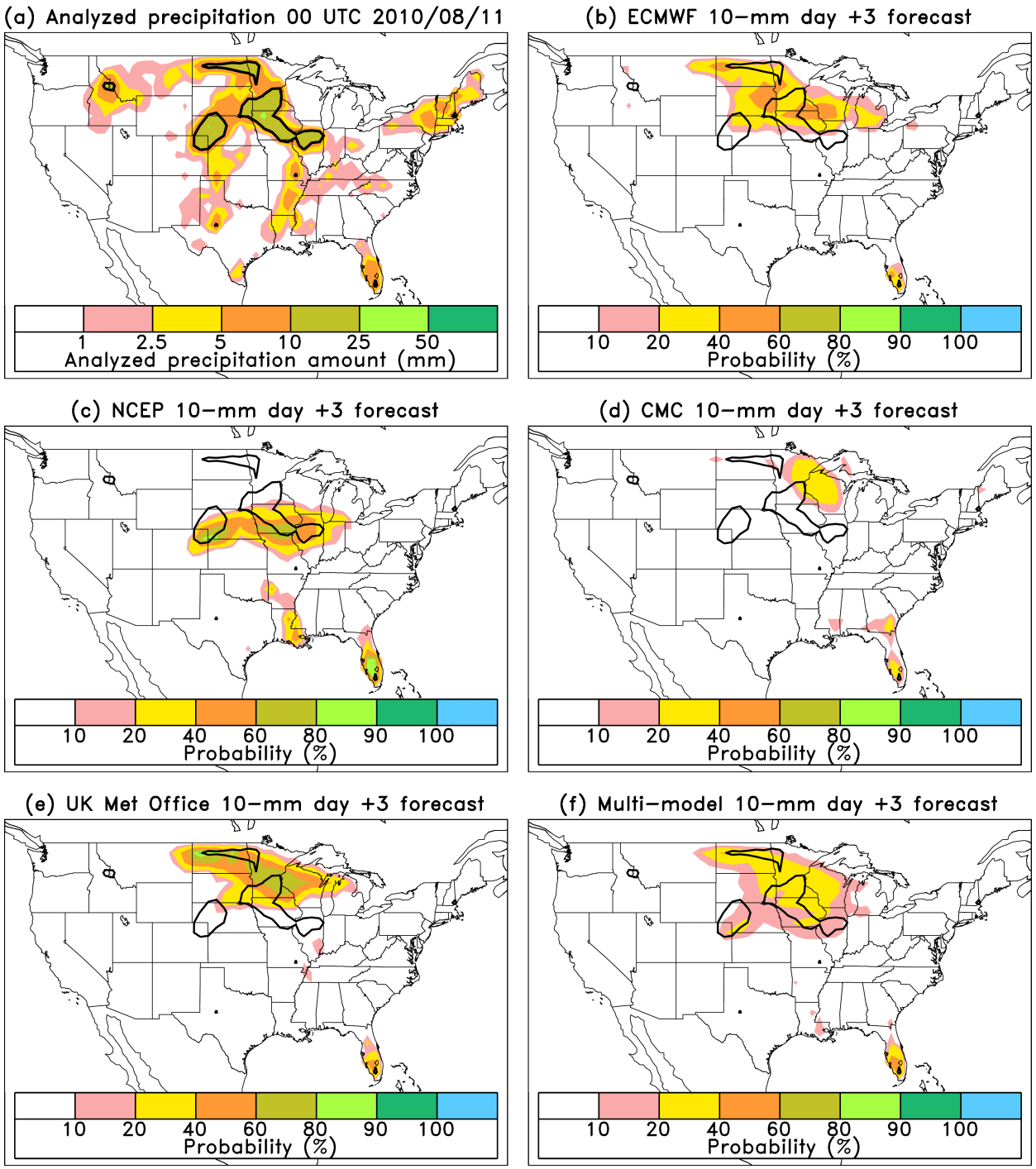
Red: > 50% forecast contour for 10-mm 24 h^{-1} area.

Multi-model & reforecast-calibration results



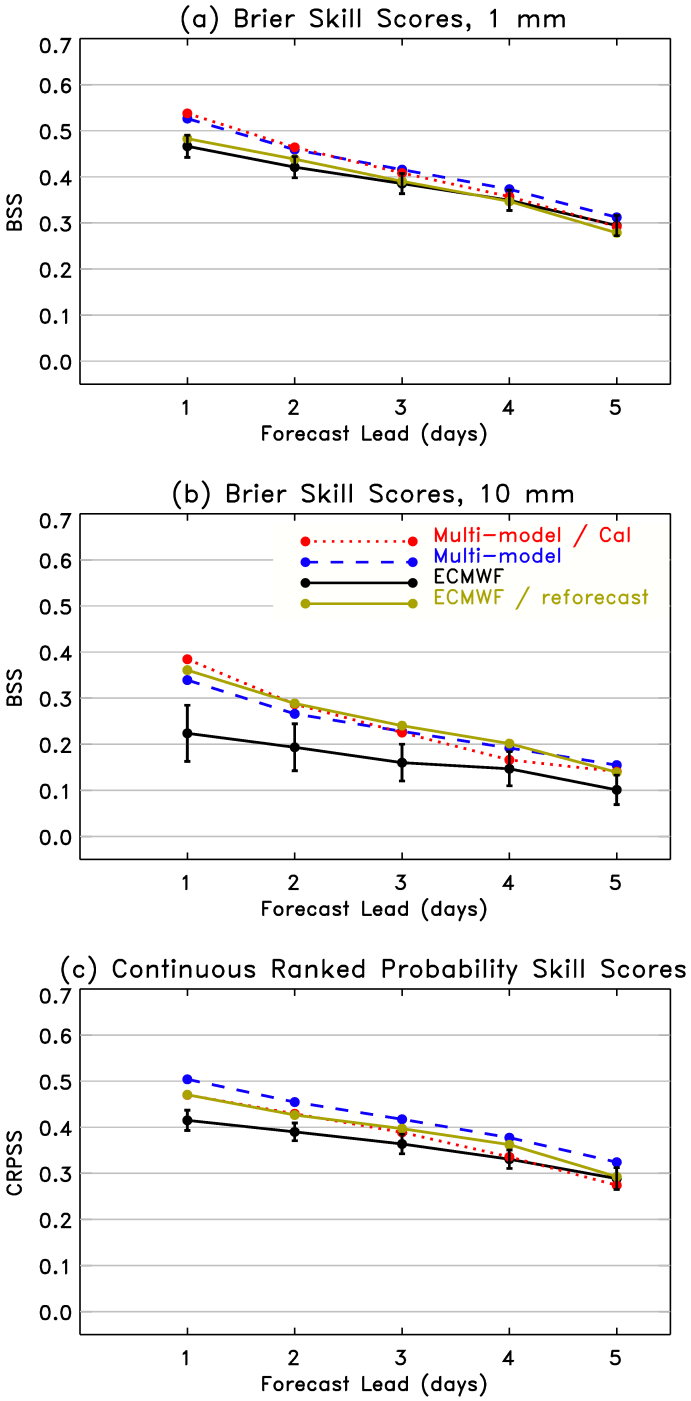
Example: where multi-model won't help.

Positional biases are similar in all the models; each is too far north.



Example: where multi-model should help.

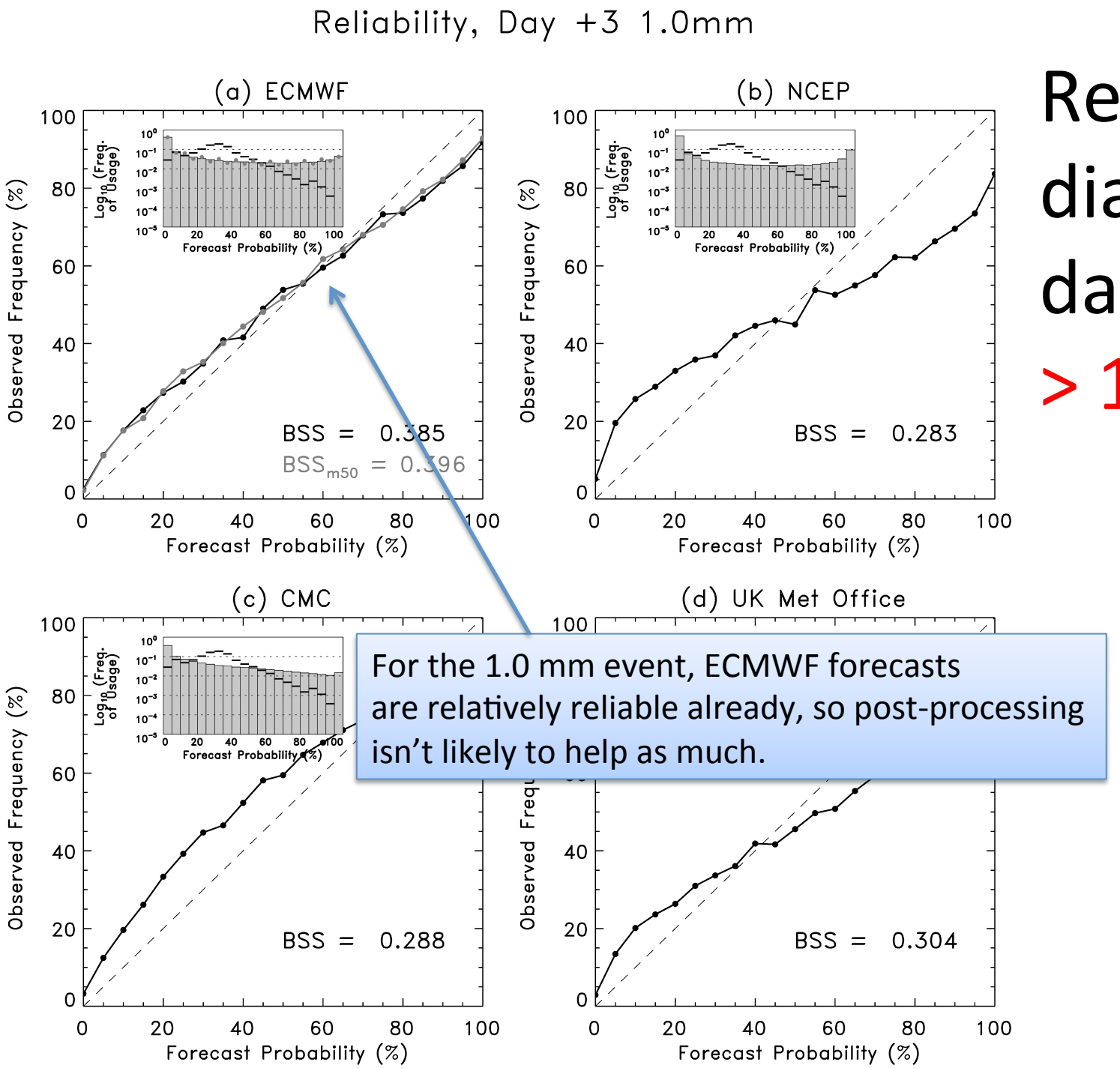
Positional biases are different; NCEP south, ECMWF north.



Skill scores for multi-model and reforecast-calibrated

Notes:

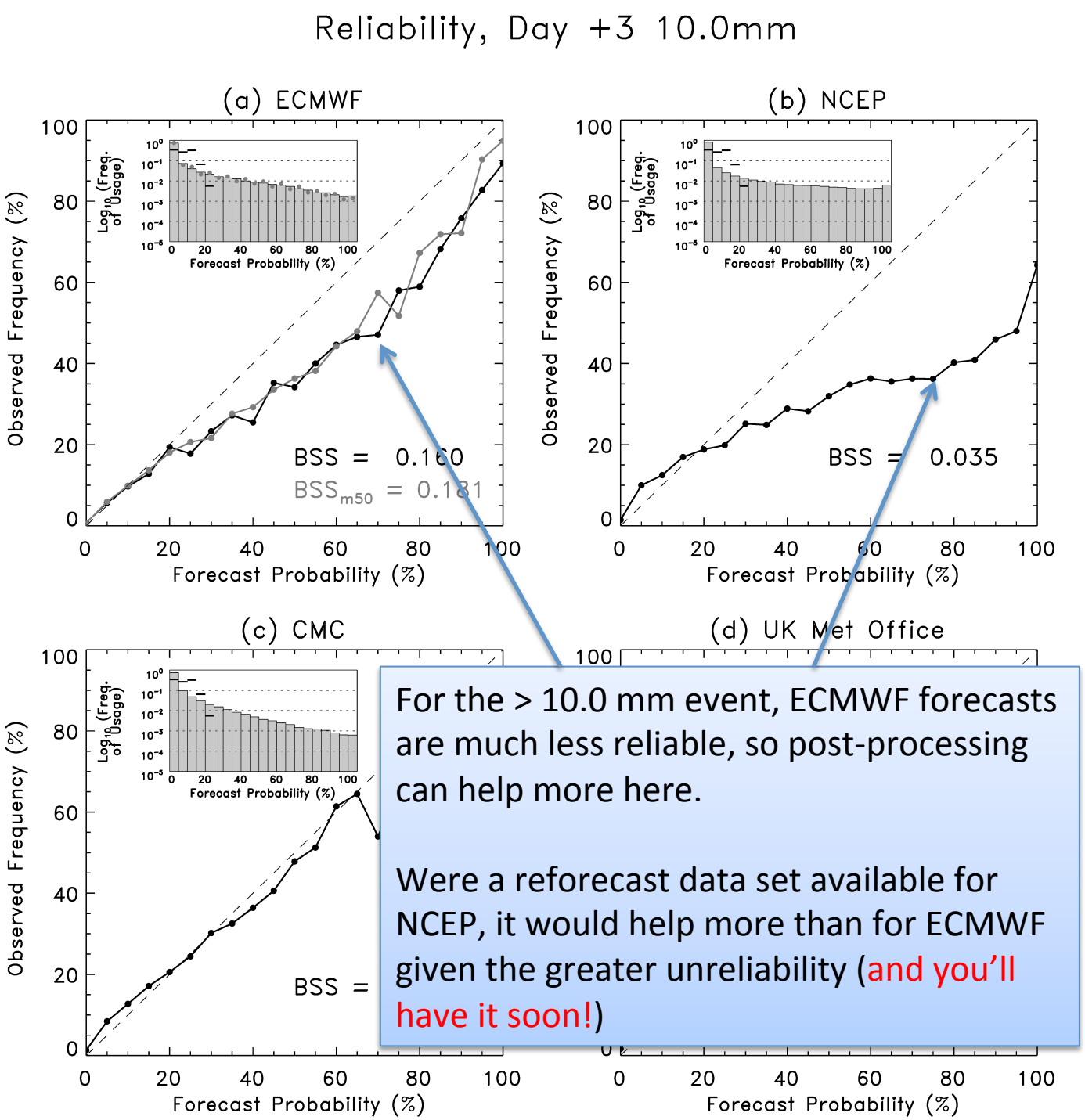
(1) Impressive skills of multi-model.
(2) Reforecast doesn't improve the 1-mm forecasts much, improves the 10-mm forecasts a lot.
(3) Calibration of multi-model using prior 30 days of forecasts doesn't add much overall.



Reliability diagrams, day +3

> 1.0 mm

For the 1.0 mm event, ECMWF forecasts are relatively reliable already, so post-processing isn't likely to help as much.



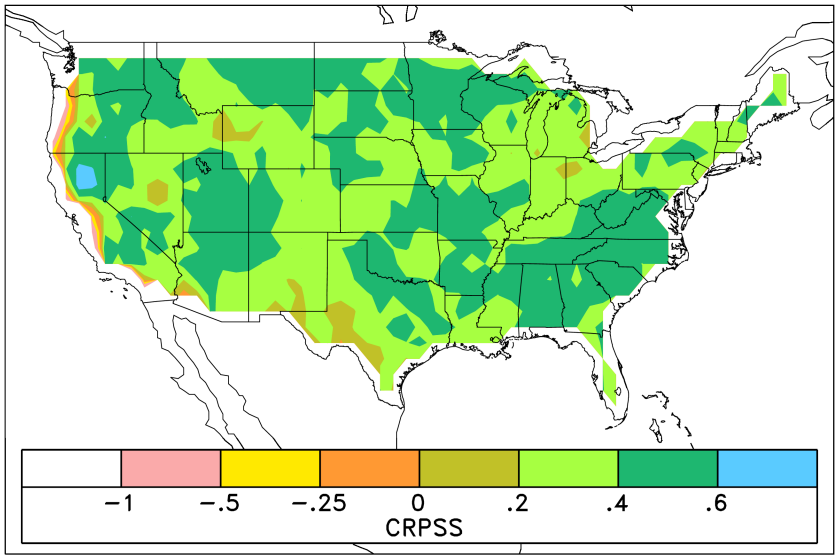
Reliability diagrams, day +3, > 10 mm

For the > 10.0 mm event, ECMWF forecasts are much less reliable, so post-processing can help more here.

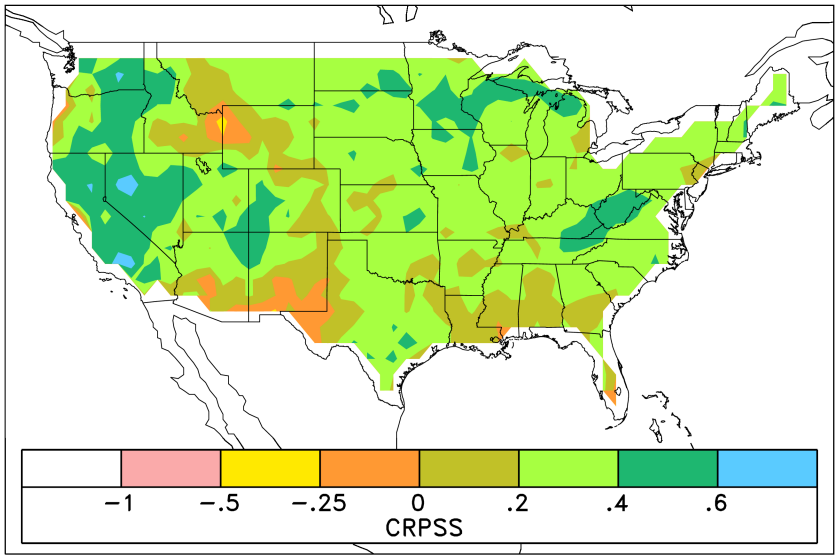
Were a reforecast data set available for NCEP, it would help more than for ECMWF given the greater unreliability (and you'll have it soon!)

CRPSS geographical distributions

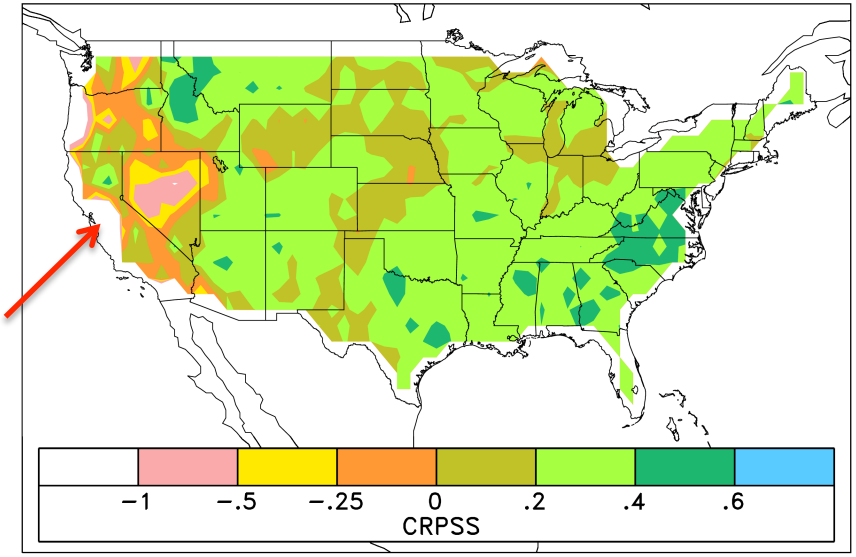
(a) ECMWF CRPSS Day +3



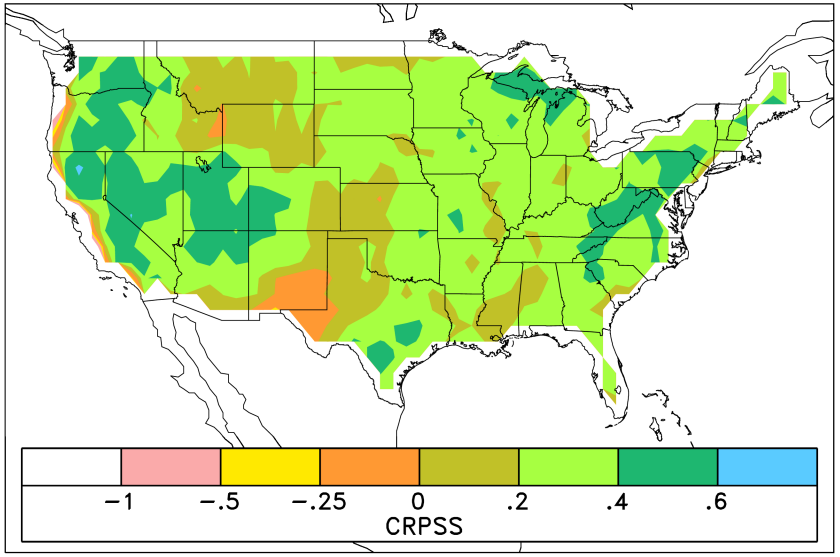
(b) NCEP CRPSS Day +3



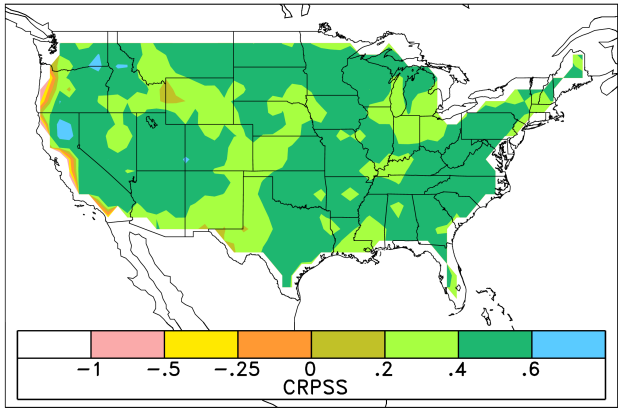
(c) UKMO CRPSS Day +3



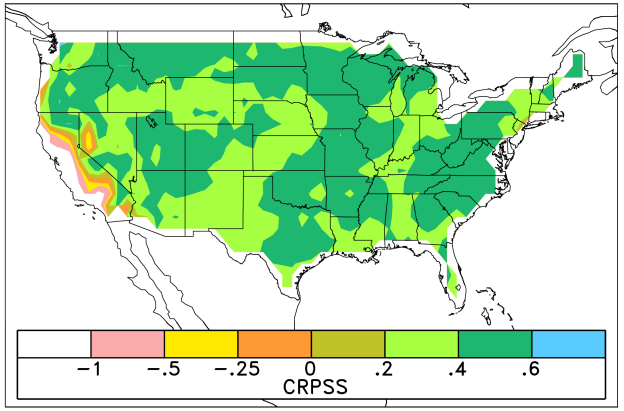
(d) CMC CRPSS Day +3



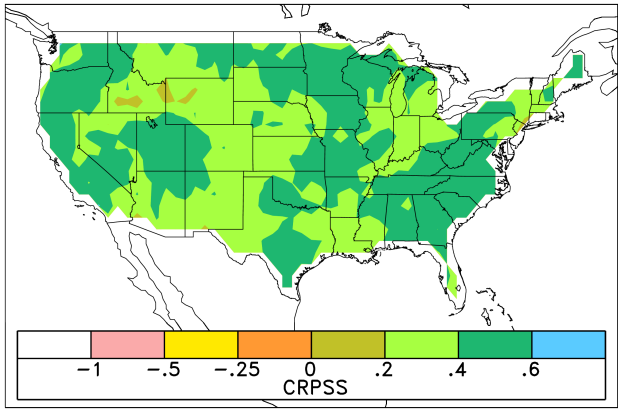
(a) Multi−model CRPSS, day +3



(b) Multi−model/calibrated, day +3



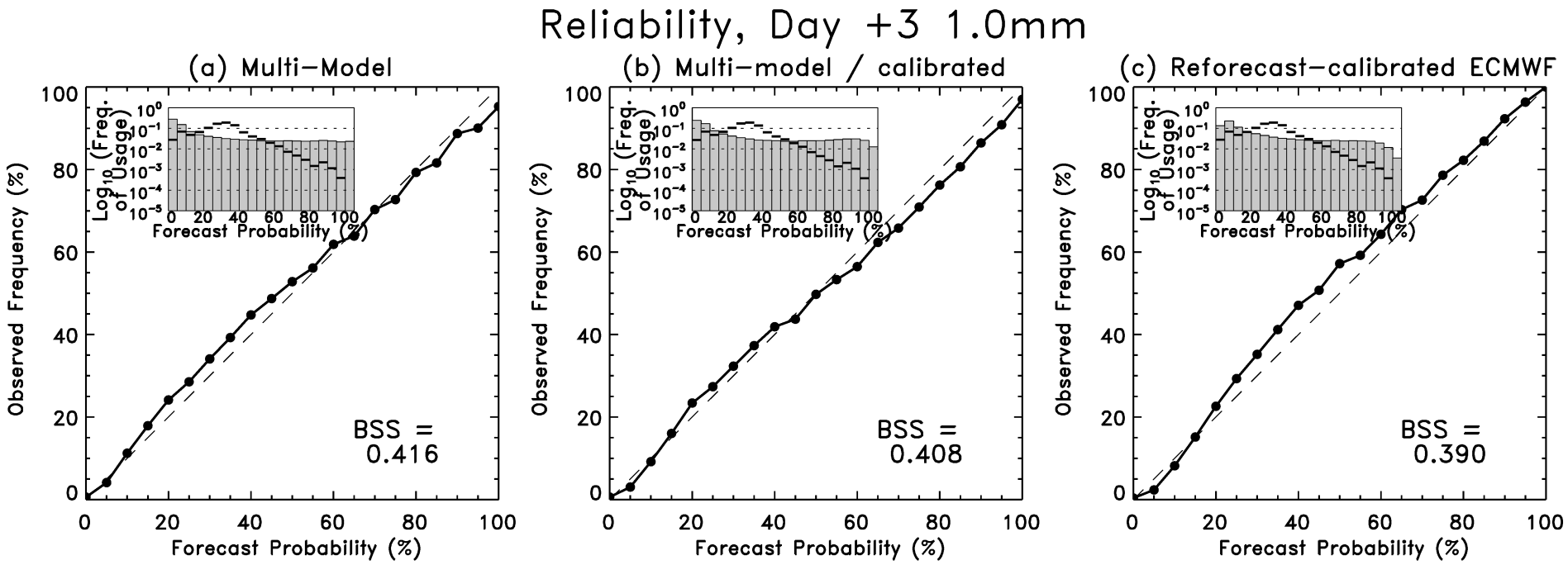
(c) ECMWF/reforecast CRPSS, day +3



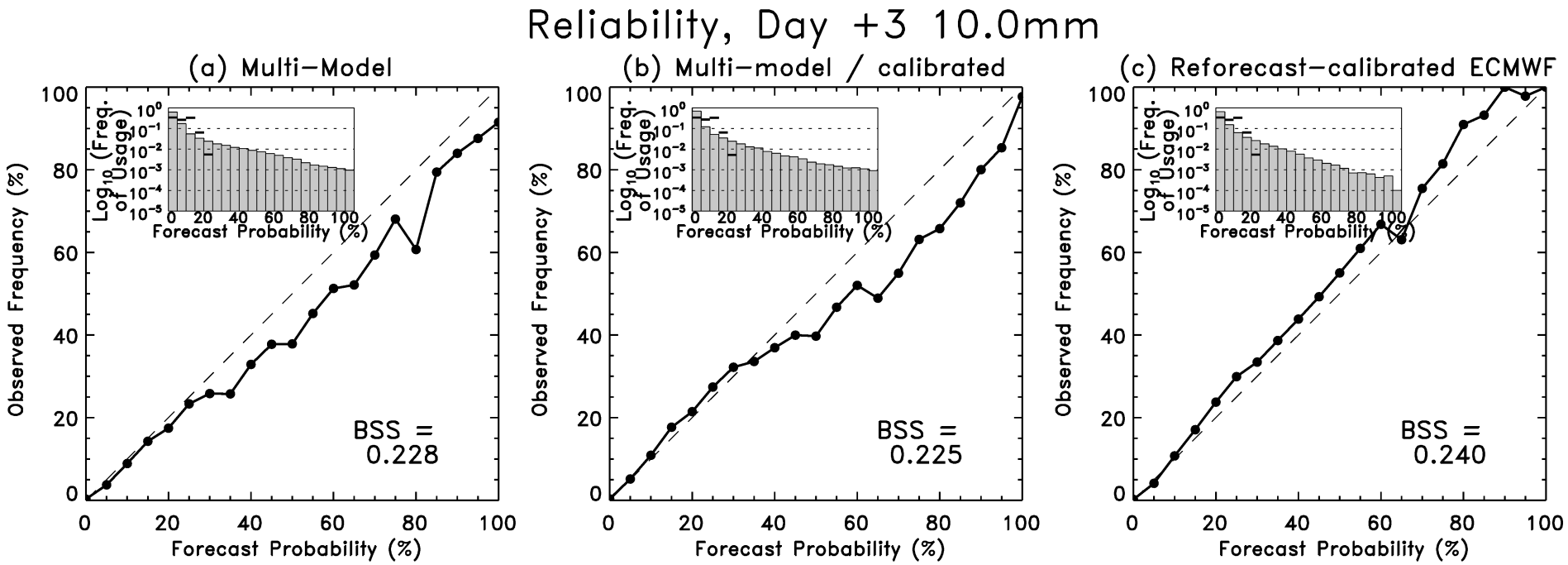
Geographic distribution of CRPSS

At this lead, multi-model calibration hurts at least as much as it helps. Small training data size.

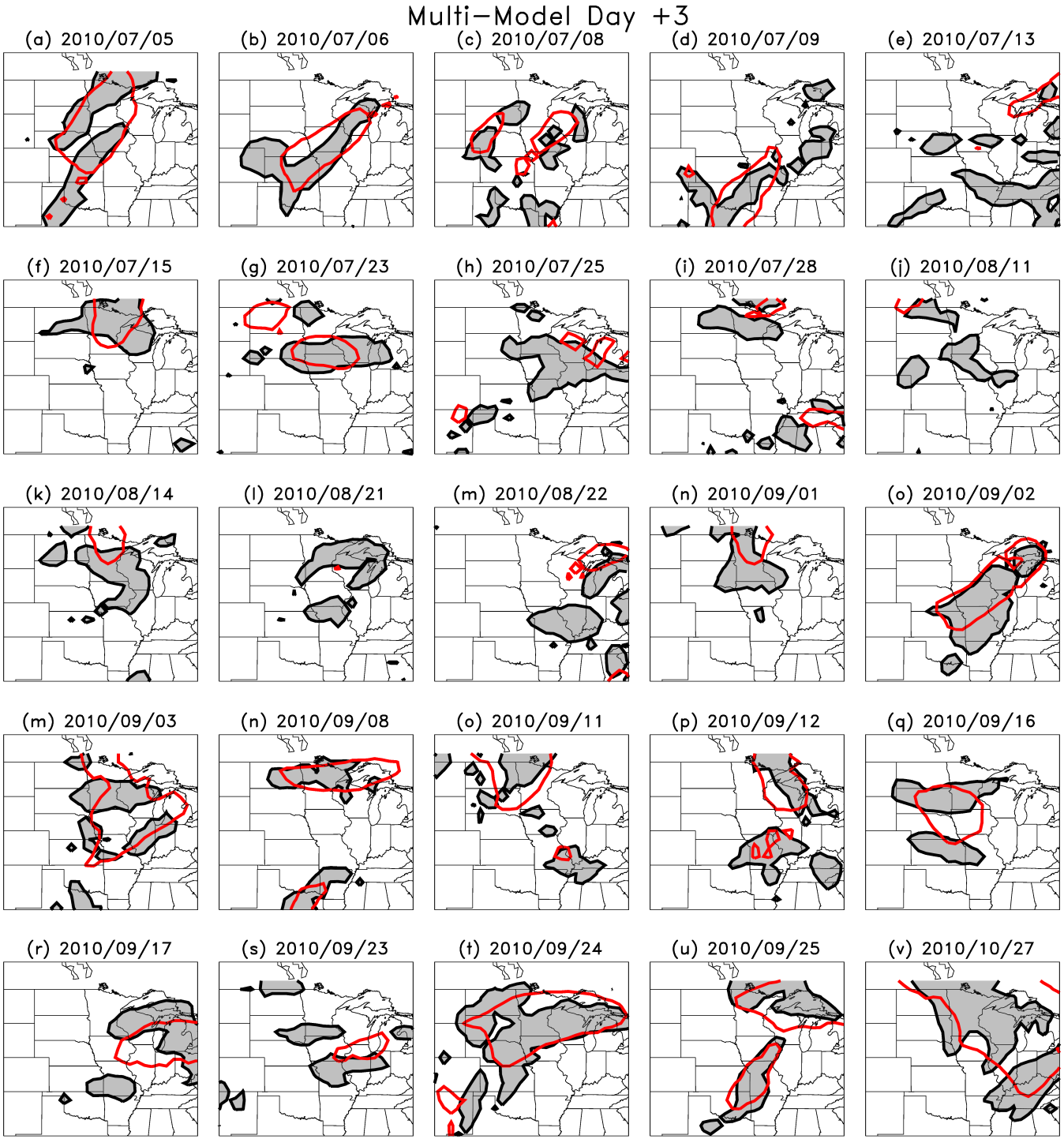
Reforecast seems to have a large impact in improving forecasts in dry areas of western US.



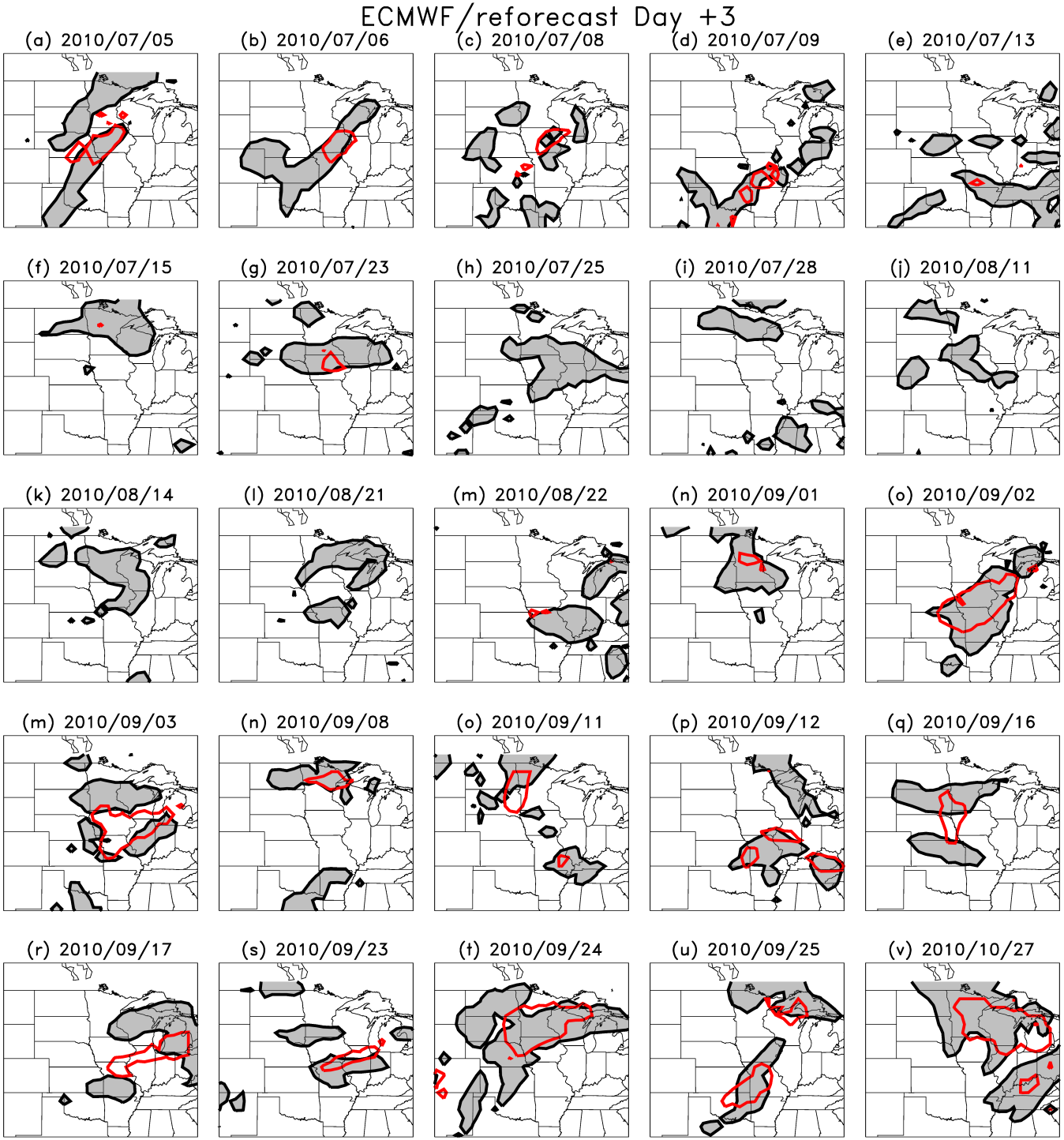
Multi-model slightly under-forecasts probabilities at 1.0 mm and is **quite reliable.** It is also substantially sharper than reforecast-calibrated, which has slightly greater under-forecast bias.



Multi-model slightly over-forecasts probabilities, and is substantially sharper. Reforecast calibrated slightly under-forecasts and is less sharp.



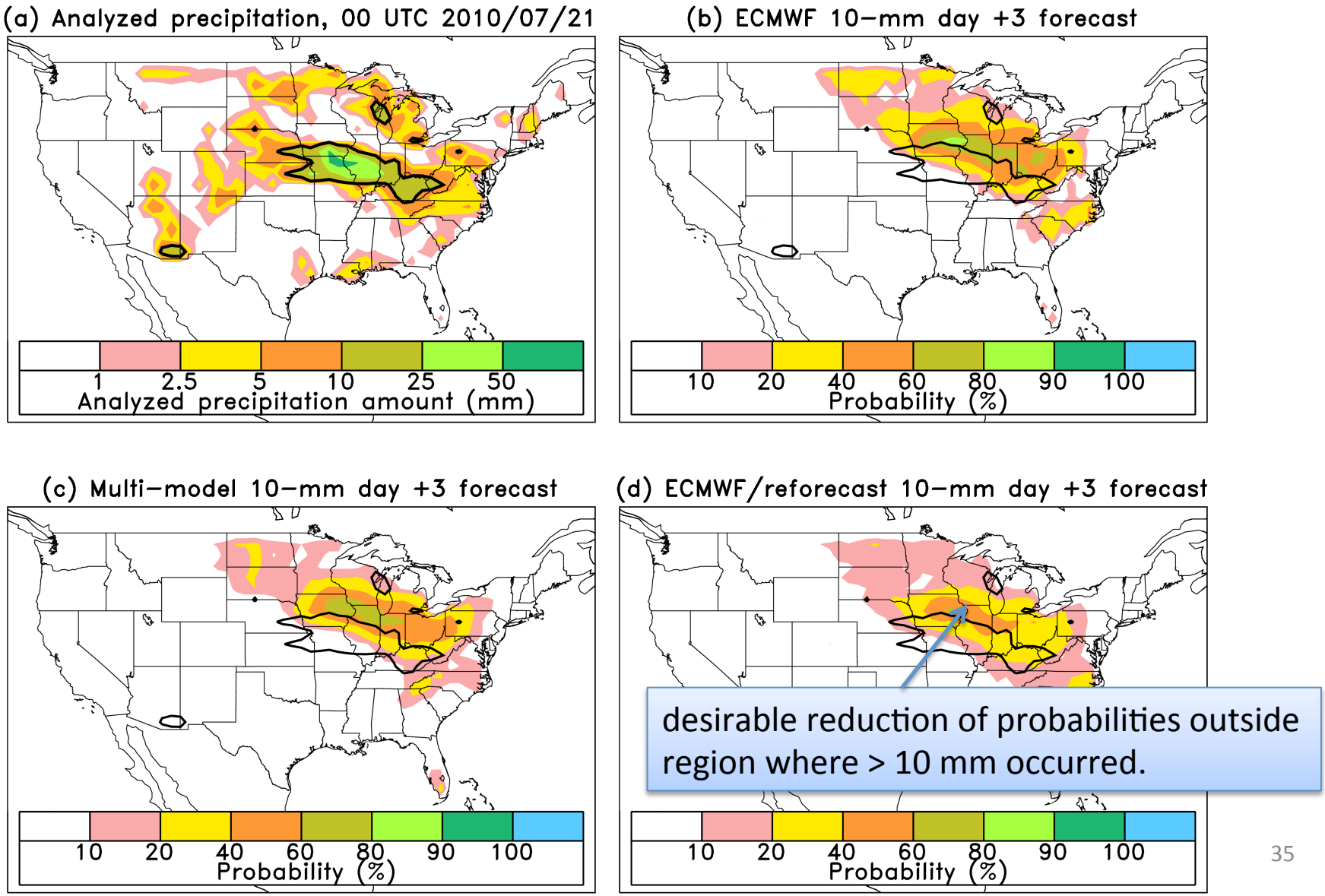
Multi-model position biases?



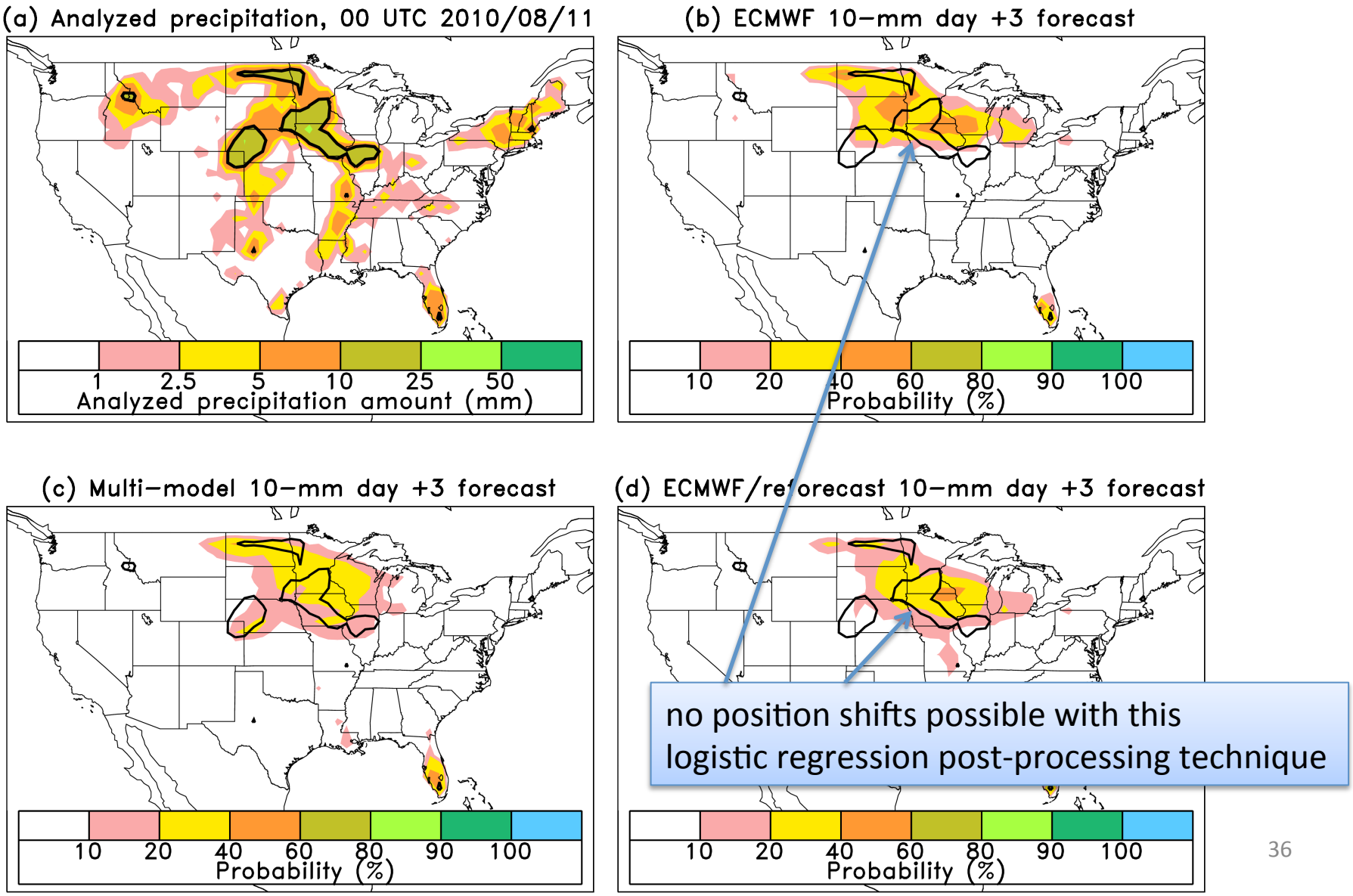
Reforecast position biases?

The most notable thing here is that the area covered by 50% is much smaller; reforecast calibration decreases sharpness.

Forecast example: 21 July 2010

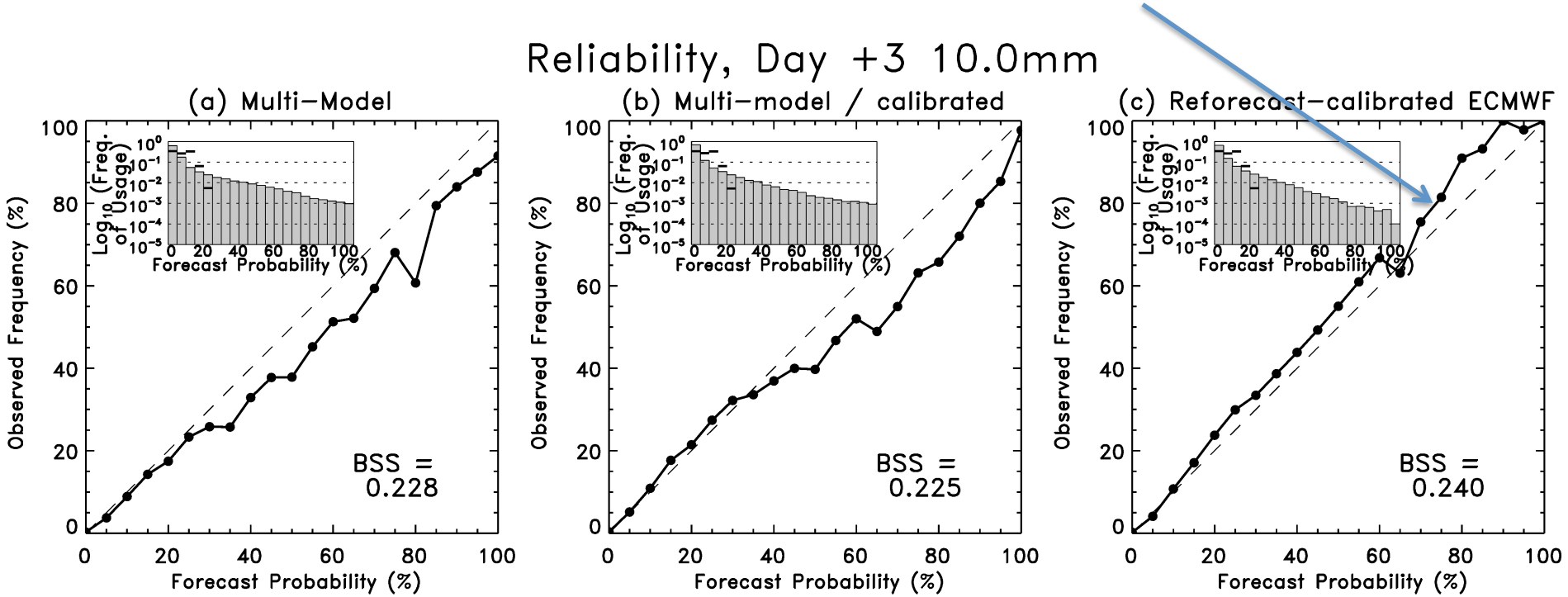


Forecast example: 11 August 2010



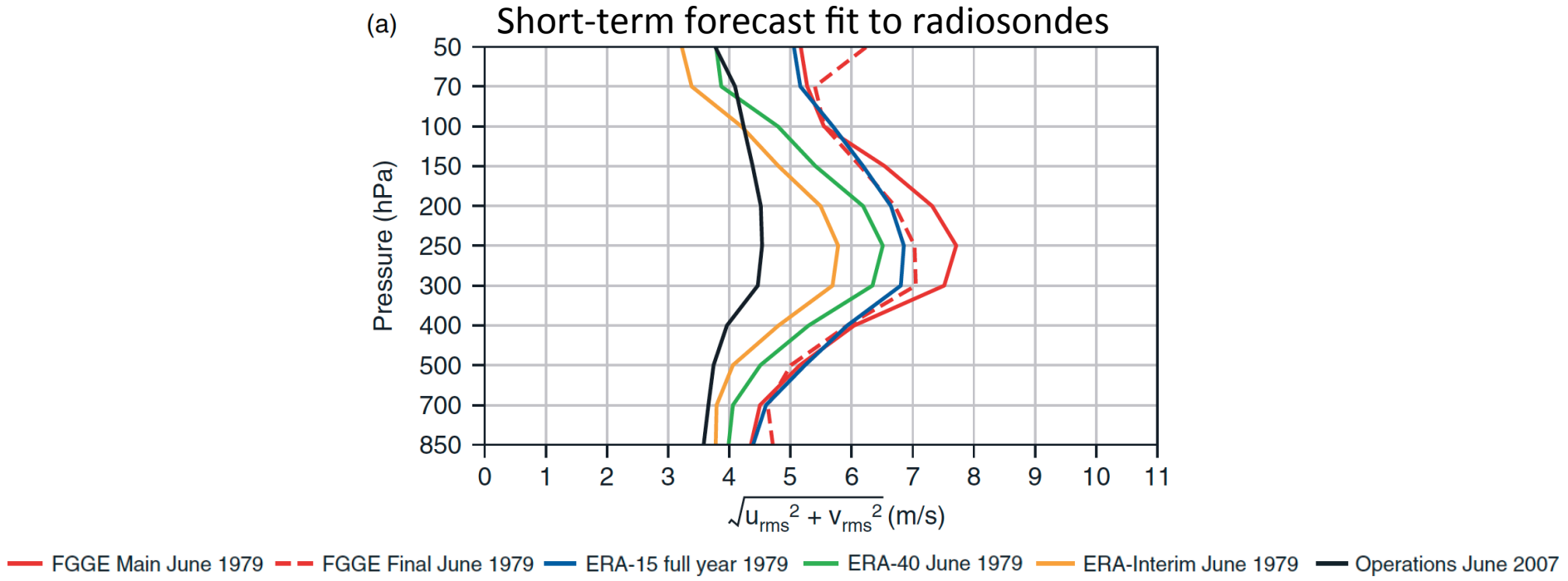
Question

- Does the assumption of consistent forecast error statistics in reforecast and real-time hold here? Is the under-forecasting here because reforecasts used in training were worse than real-time forecasts?



Accuracy of short-term forecasts from various ECMWF analyses

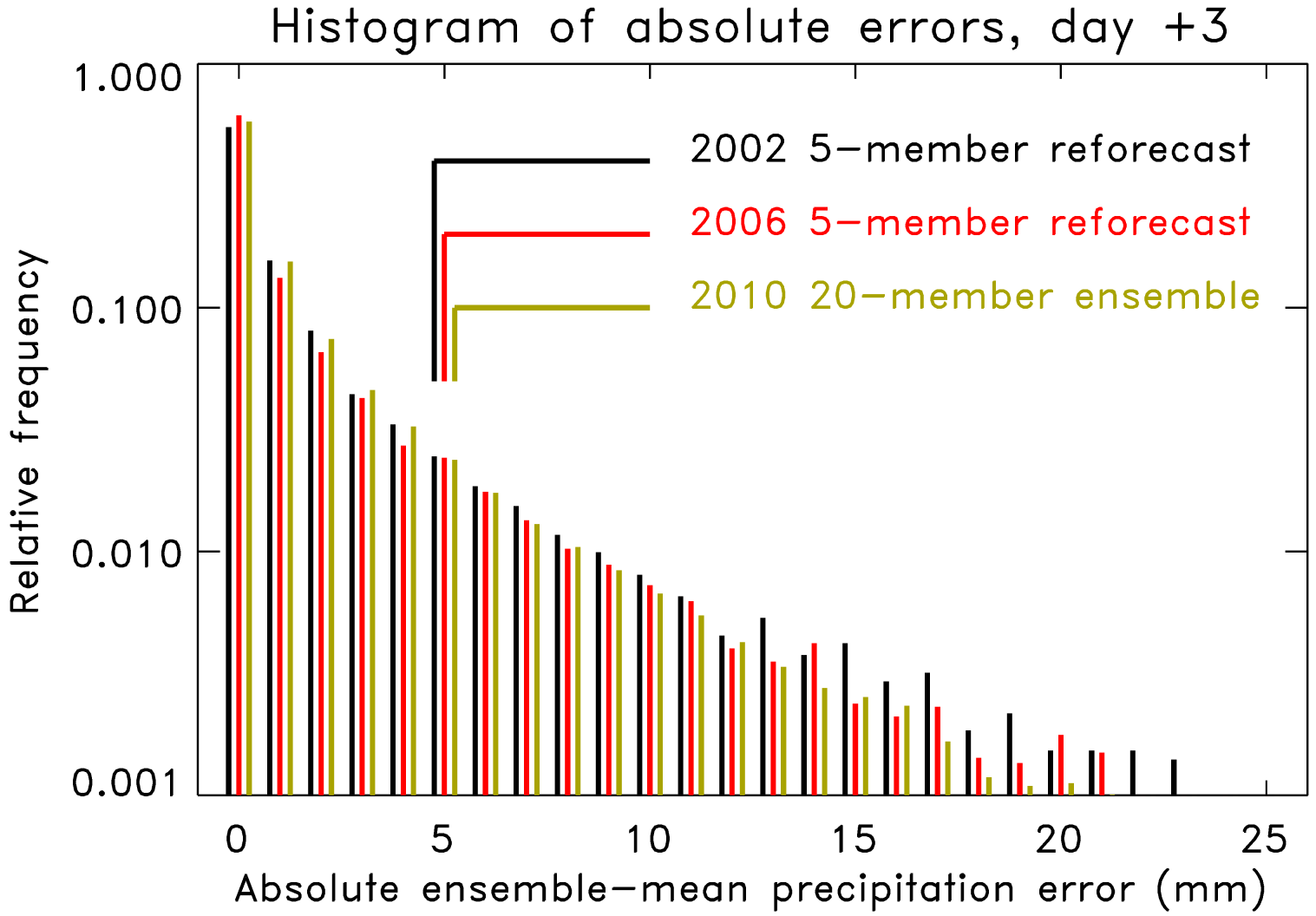
- If real climate or model-error statistics change significantly during reforecast period, decreased accuracy of post-processed estimates. Here, forecast error in past larger than for real-time forecasts.



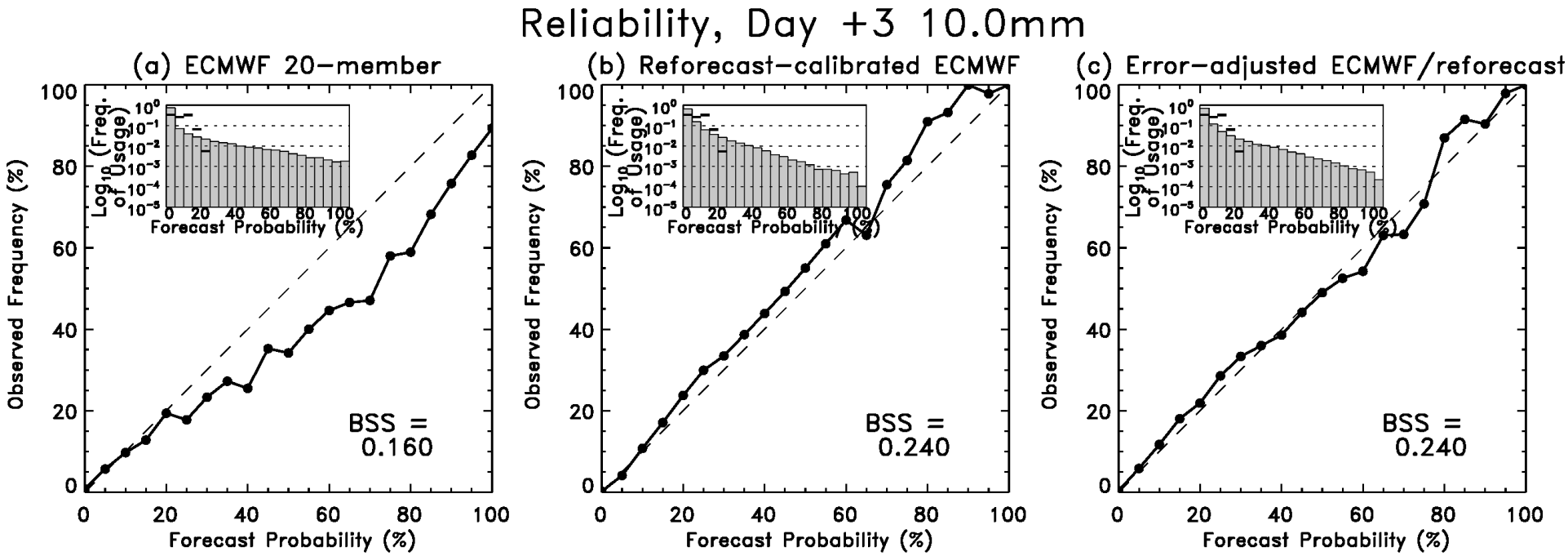
From Dee et al., QJRMS, 2011 article on ERA-Interim

38

Are 2010 ECMWF real-time forecasts more accurate than older reforecasts?



Reforecast calibration after training data relaxed slightly toward analyses



Here, for the older reforecasts, the reforecast training data was (arbitrarily) nudged slightly toward the analysis, to simulate 2010 data. 10 % nudging for 2002 data, 7% nudging for 2009 data.

Result: slightly more reliable & less sharp at highest probability, but no overall gain in skill.

Change of topics: update on GEFS reforecasts

Recall we're doing a 30-year, once-daily, 11 member reforecast using 2012 GEFS system and saving a lot more of the data than we did for the reforecast with the 1998 model.

GEFS reforecast status report

- Control
  - all 00Z reforecasts 1979-2009 done to 16 days lead.
  - all 12Z reforecasts 1979-2009 done to 8 days lead.
- Perturbed initial conditions generated (ETR)
- 11 years of week-1 reforecasts complete; estimate complete by late Nov 2011
- 0 years of week-2 reforecasts complete; estimate complete by early 2012
- 170 TB archival system in place at ESRL
  - all 00Z control runs transferred to ESRL
  - ~3 years of the 10-member ensemble transferred to ESRL
- Just starting development of software to serve out the reforecast data conveniently to you (http, ftp, openDAP).

Conclusions

- Hypothesis not confirmed; as opposed to T_{sfc}, multi-model slightly better than reforecast, except for heavier precipitation.
  - Reforecast more reliable, multi-model sharper.
- Reforecast limitations here:
  - Post-processing based on 2002-2009 data only, constrained by observational availability.
  - ECMWF's reforecast data set shown to be non-homogeneous, with larger errors in past than for real-time data.
  - More complicated post-processing techniques not tried (yet).
- Gain in skill from post-processing ECMWF with reforecasts much larger than gain in multi-model using short training data set, especially for 10-mm forecasts. Illustration of the power of the large sample size that reforecasts afford.
- Will push, via THORPEX, for more real-time data sharing.
- GEFS reforecast available soon, consistent with your 2012 configuration.

Acknowledgments

- Baudouin Raoult, ECMWF, for TIGGE web site interface & help.
- Florian Pappenberger, ECMWF, for reforecast data.
- Roberto Buizza, ECMWF, for constructive criticism.
- Yan Luo of EMC for CCPA support.

Verification details

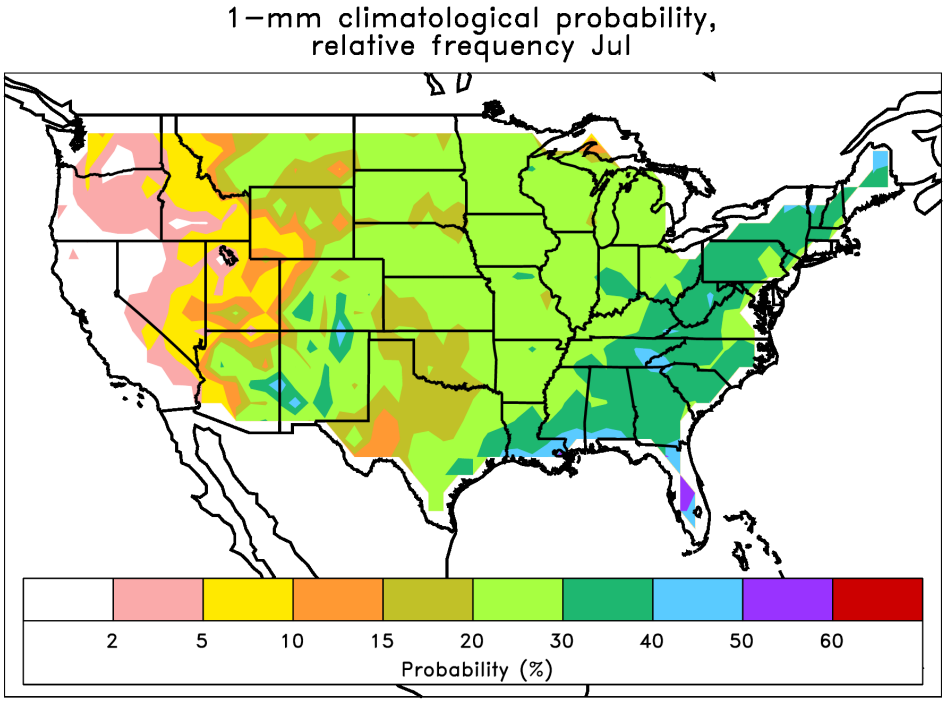
- Brier skill scores
- CRPSS

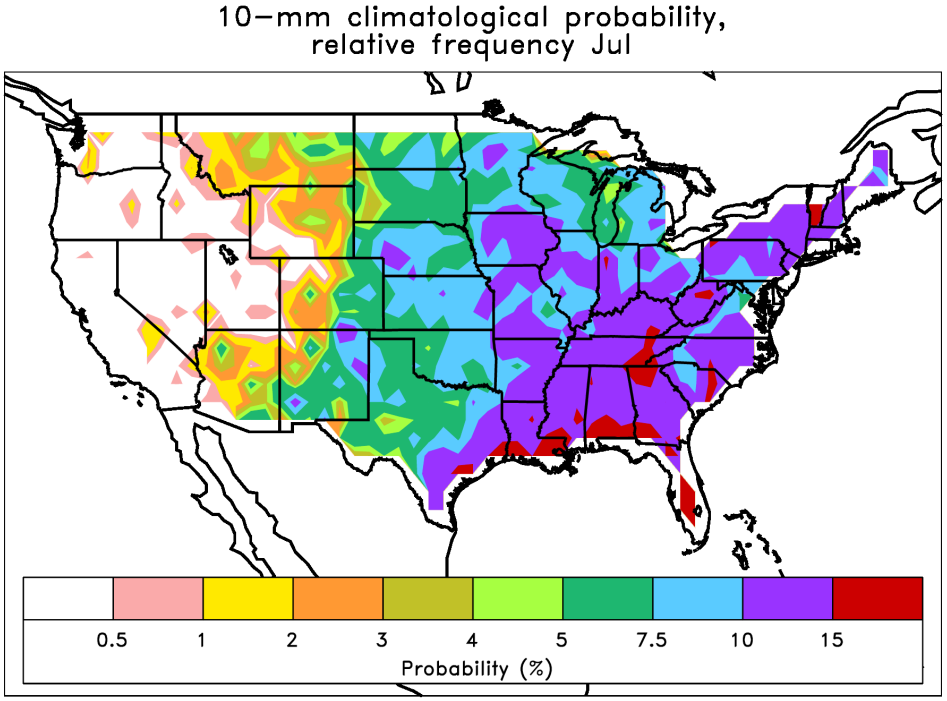
[go back]

Calculation of climatological probabilities for BSS reference

- Based on climatology of EMC CCPA product, 2002-2009. 1-degree analyses
  - CCPA attempts to make the Stage-IV radar/obs blended product look similar statistically to purely gage product, via regression analysis.
  - In some dry areas CCPA (regression-based) approach does not have enough data to work properly. There, replace with Stage-IV data.

- Climatological probabilities determined separately for each month.

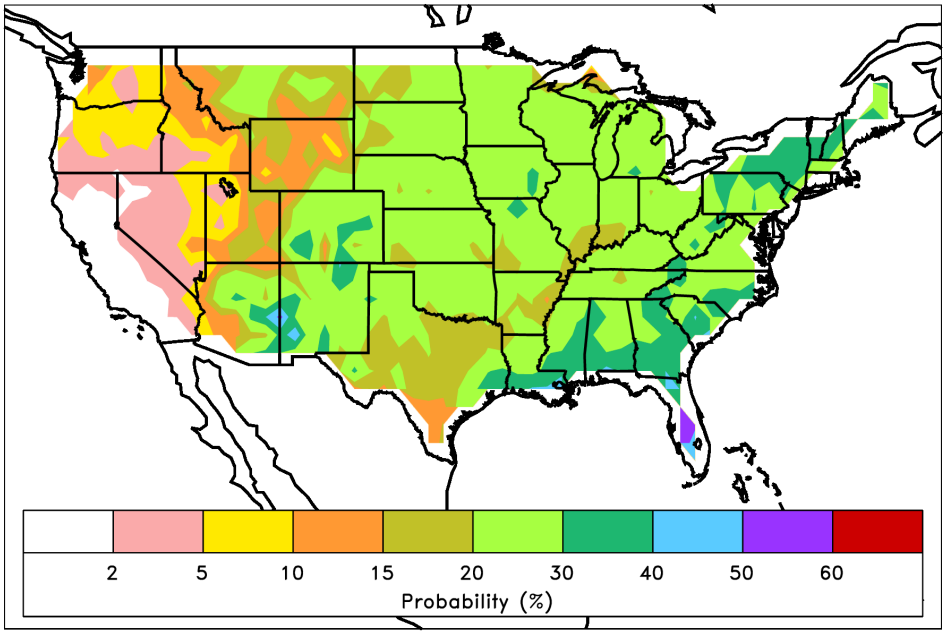




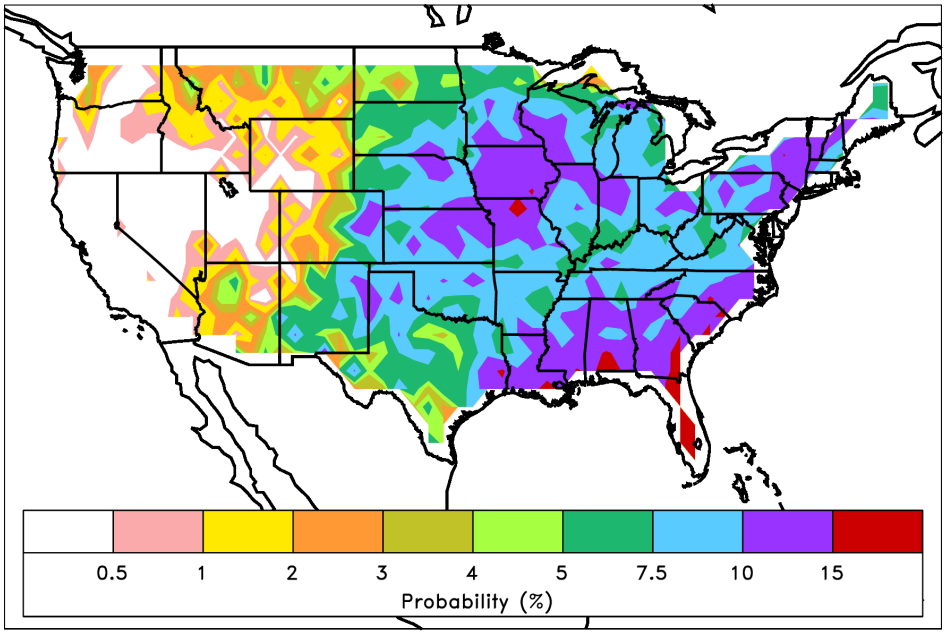
Climatological probability of
> 1 mm/24h and 10mm/24h

July

1−mm climatological probability,
relative frequency Aug

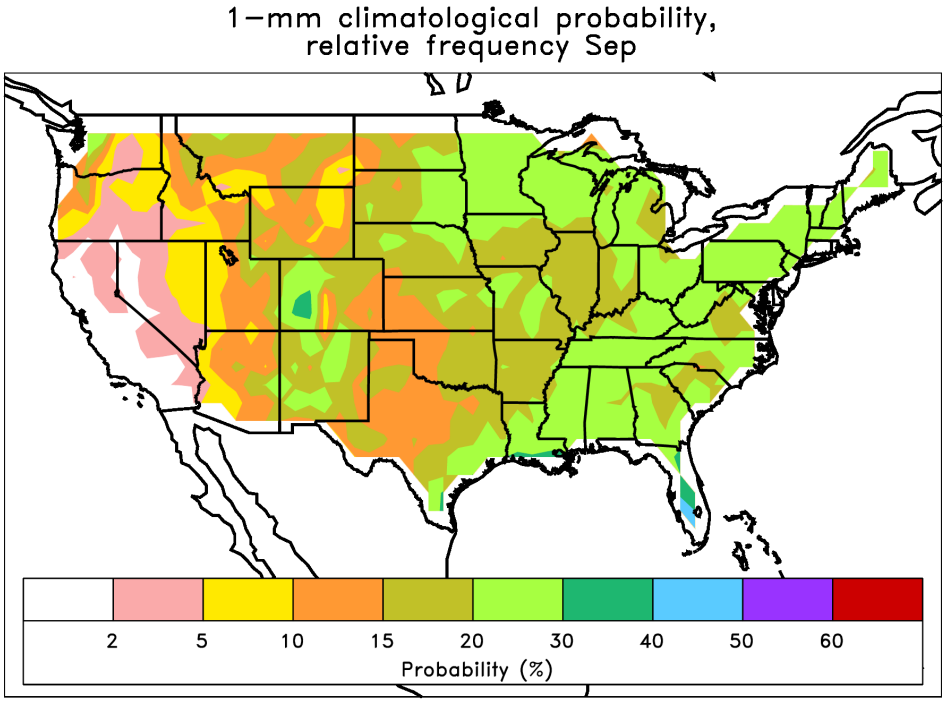


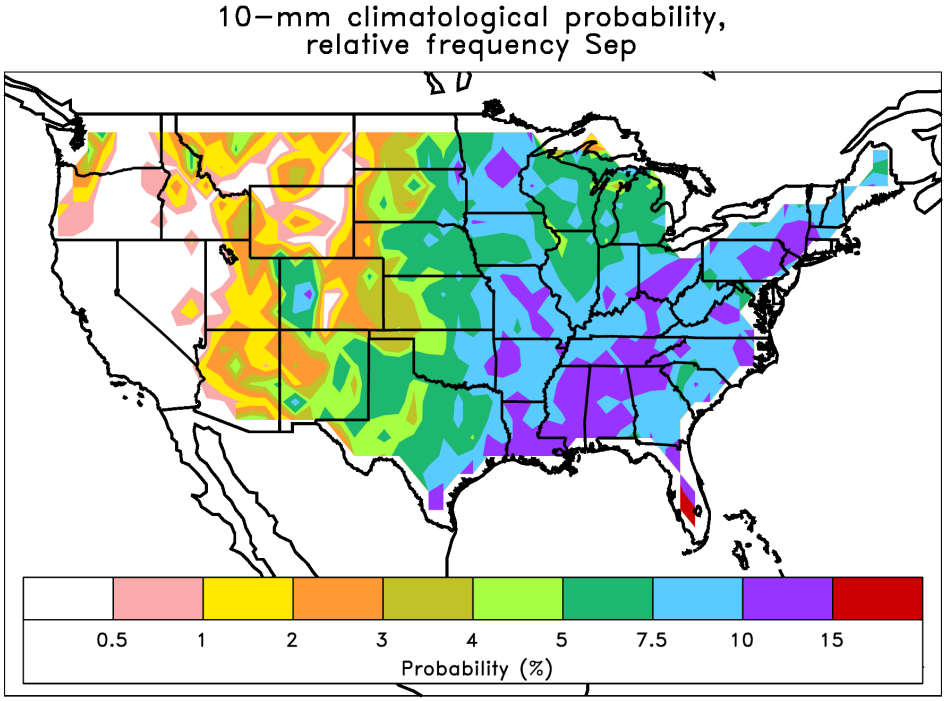
10−mm climatological probability,
relative frequency Aug



Climatological probability of
> 1 mm/24h and 10mm/24h

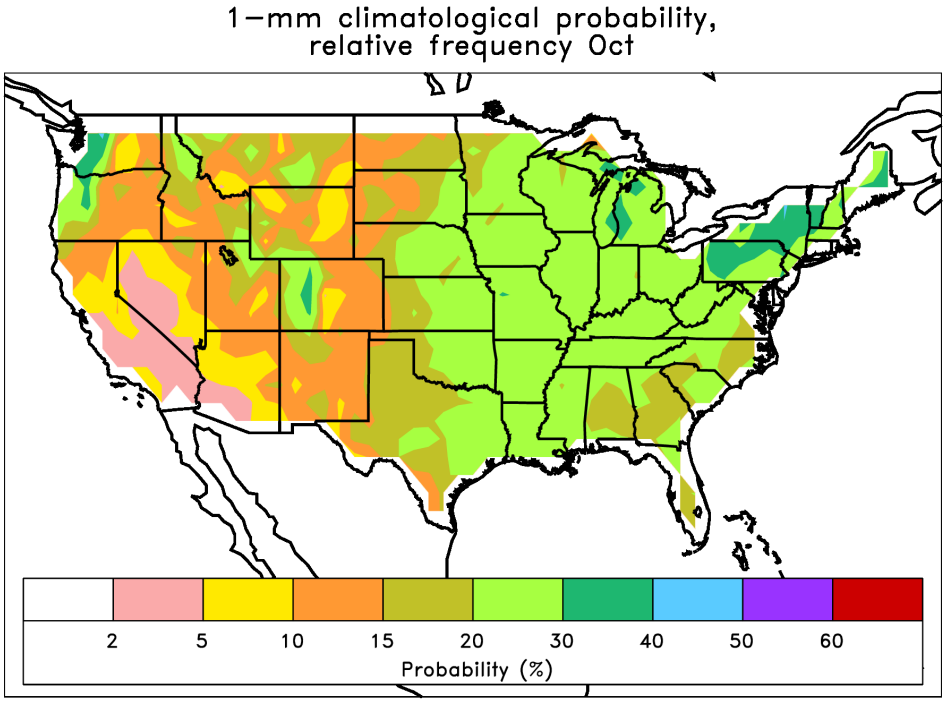
August

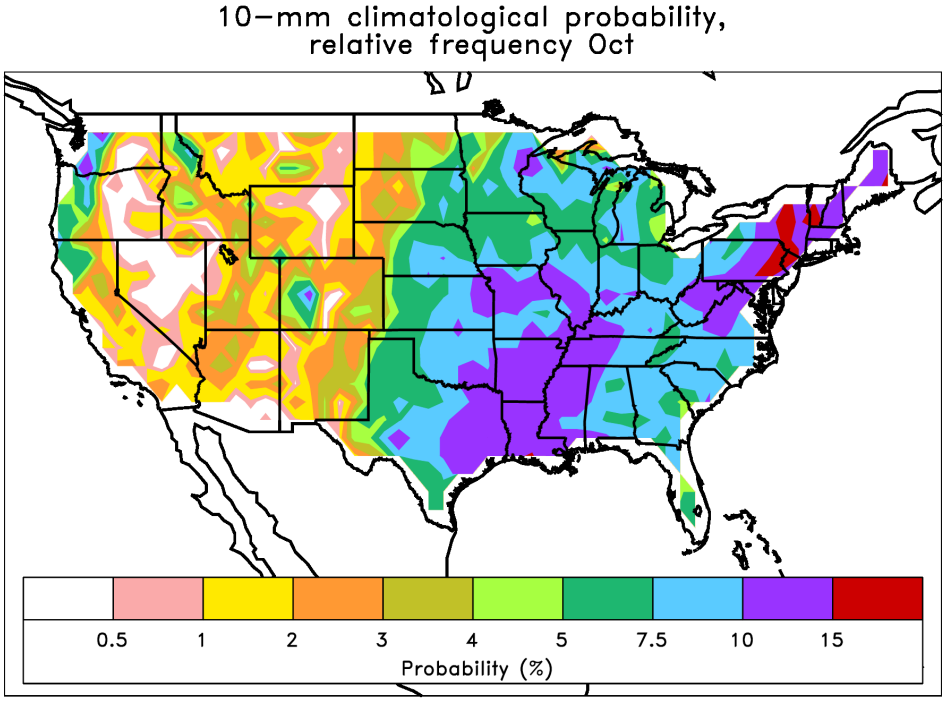




Climatological probability of
> 1 mm/24h and 10mm/24h

September





Climatological probability of
> 1 mm/24h and 10mm/24h

October

Brier score and skill score (conventional method)

$$BS_f = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - o_i\right)^2$$

$$BSS = 1.0 - \frac{BS_f}{BS_{cl}}$$

My past research has discussed how this can over-estimate forecast skill (Hamill and Juras, October 2006 QJRMS).

My calculation of Brier skill scores

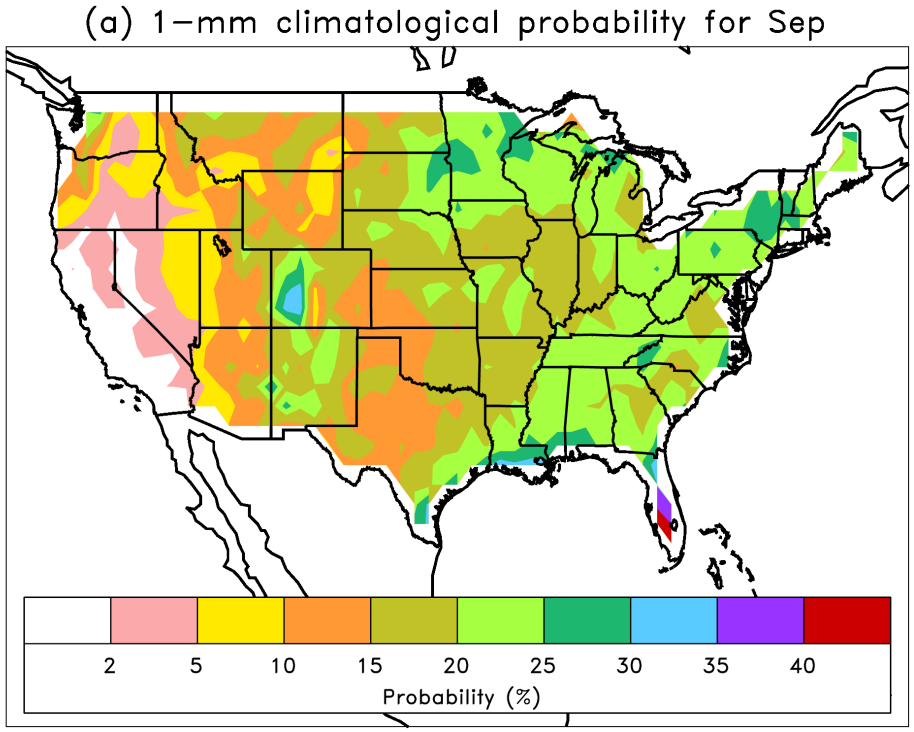
General idea is to compute *BSS* as average of *BSS* over a set of locations/times ("classes") that have more similar climatological probabilities. This minimizes problem of over-forecasting skill. Here I use 6 classes.

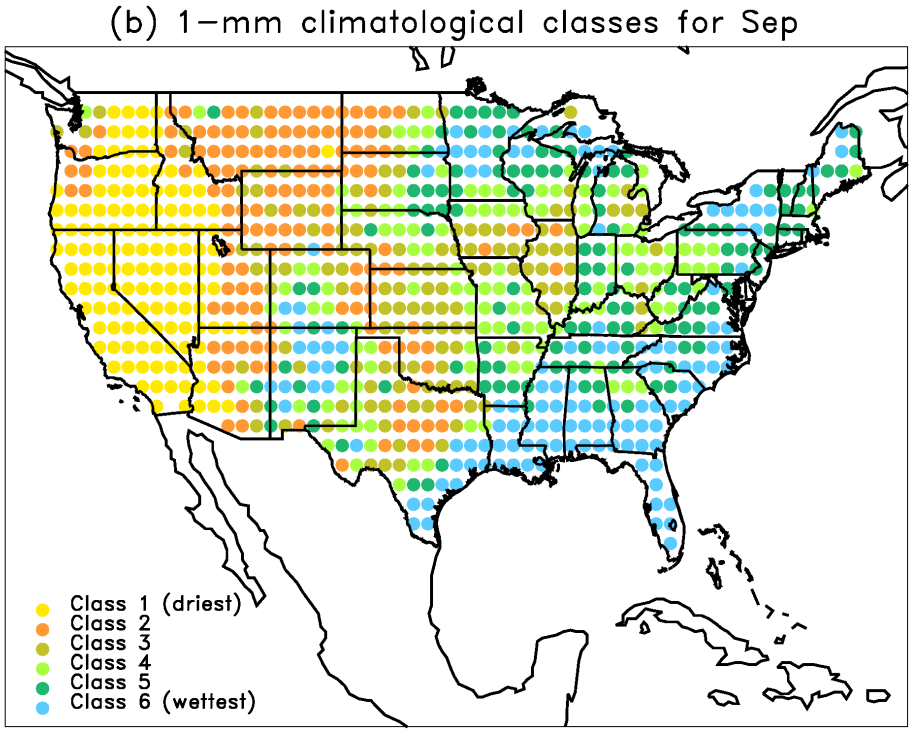
$\mathbf{BS}^{f1} = \left[\mathbf{bs}_1^{f1}, \ldots, \mathbf{bs}_6^{f1} \right]$ matrix of Brier scores for forecast model *f1*, where $\mathbf{bs}_i^{f1}$ was a n_d –dimensional (= 123, the number of case days here) column vector of average Brier scores for the points in the i^{th} class and for forecast model *f1*.

$\overline{\mathbf{bs}}^{f1} = \left[\overline{bs}_1^{f1}, \ldots, \overline{bs}_6^{f1} \right]$ The average over the 123 case days

$$BSS = \sum_{k=1}^{6} \frac{1}{6} \left(1 - \frac{\overline{bs}_k^{f1}}{\overline{\overline{bs}}_k^{c}} \right)$$

Climatological probabilities and class





Computation of CRPSS

$s = 1, \ldots, n_d \times n_s$ samples (# days * # grid pts)

$\mathbf{q}_s = \left[q_1^s, \ldots, q_{20}^s \right]$ be the 20-dimensional vector of the precipitation quantiles associated with the 2.5th, 7.5th, …, 97.5th percentiles of the climatological *CDF* for that point and that month for the s^{th} sample.

$$CRPS_f = \frac{\sum_{s=1}^{n_d \times n_s} \cos(\phi_s) \sum_{iq=1}^{20} 0.05 \times \left[F^s\left(q_{iq}^s\right) - O^s\left(q_{iq}^s\right) \right]^2}{\sum_{s=1}^{n_d \times n_s} \cos(\phi_s)}$$

$F^s\left(q_{iq}^s\right)$ represents the forecast's *CDF* for the s^{th} sample evaluated at the q_{iq}^s quantile

$O^s\left(q_{iq}^s\right)$ same but for observed

ϕ_s is the latitude of the grid box

$CRPSS = 1. - CRPS_f / CRPS_c$